

MATHEMATICAL INSTITUTE

MASTER THESIS

STATISTICAL SCIENCE FOR THE LIFE AND BEHAVIOURAL SCIENCES

---

# Causal inference on chemotherapy modifications

---

Author:  
Katerina Bakunina

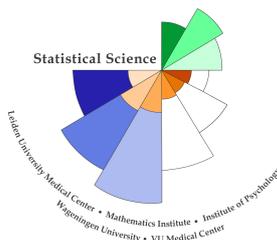
First Supervisor:  
Dr. Marta Fiocco  
LUMC & LU

Second Supervisor:  
Dr. Carlo Lancia  
LUMC & LU

November 2016



**Universiteit  
Leiden**  
The Netherlands



LEIDEN UNIVERSITY MEDICAL CENTER

## Abstract

Nowadays data from randomised experiments is used to assess relationships beyond the primary goal of the study. From the perspective of secondary analysis, the exposure of interest is no longer allocated at random to the experimental units, and the existence of confounders is almost certain.

Randomised clinical trials usually have longitudinal nature. In this case the confounding may be time-dependent, i.e. the value of the confounder (as well as the exposure) may vary over time. Furthermore, the exposure often influences future values of the confounders. This two-way relationship is referred to as *exposure-confounder feedback*.

In this thesis we use data from a randomised controlled trial for chemotherapy in osteosarcoma to illustrate the methodology for causal inference in the presence of time-dependent confounding and exposure-confounder feedback. We build Marginal Structural Models (MSMs) for binary and time-to-event outcomes, and use inverse-probability-of-treatment weighted (IPTW) estimation method.

The thesis novelty is twofold. First, we illustrate how to build MSMs by using chemotherapy data. Second, we discuss how to simultaneously assess the causal effects of time-varying and point exposures.

The study findings indicate that closer collaboration between oncologists and surgeons is required when treating osteosarcoma patients, since surgery delay has a strong negative effect on patients' histological response (HRe) (an intermediate outcome, which indicates chemotherapy effectiveness). Although the data provides some evidence for a weak protective effect of surgery delay on the hazard of death, there is indication that surgery delay has much stronger negative effect on the hazard of disease progression and/or cancer recurrence.

Based on the results from the analyses, revision of the chemotherapy drugs dosage might be discussed in the clinical community. We found that smaller doses are associated with good HRe and decrease the chance of any long-term adverse event.

## Acknowledgements

I wish to express my deepest regards to my internship and thesis supervisor Dr. Marta Fiocco. First of all, for providing me with the opportunity to work under her supervision on two challenging and demanding projects. Second, for the freedom and independence that allowed me to discover how much the master programme *Statistical Science for Life and Behavioural Sciences* had taught me, and how capable I have become. Her comments were critical, encouraging as well as fruitful, nevertheless the one year of our joint contribution to the development of healthcare science felt more like a collaboration than working under supervision. And when it is finally over, I can say: I would not wish it to be any other way.

My second thesis supervisor, Dr. Carlo Lancia, deserves no less gratitude. Thank you for always reaching out and all the dozens of hours that you spent on our discussions and reading my thesis drafts.

I wish to acknowledge Dr. Jakob Anninga for providing the research team with understanding of the clinical background of the problem under study and the processes that we were modelling.

I wish to thank all contributors of the master track *Statistical Science for Life and Behavioural Sciences*, from its design to execution. Being a student in this master programme was the best time of my life combining challenges with interesting discussions and a lot of fun. Among my fellow students I found more adherents than ever before.

And last but not least, I wish to thank my family and closest members of my personal life. First, to my mother for providing me with this life-changing opportunity to follow the statistical science master programme in Leiden University. Second, to my wife for sharing with me the moments of achievement as well as cheering me up when my spirit was low. Her second-person opinion allowed me to learn more from the feedback that I received, than I could have done myself alone. I wish to thank all awesome people that I met in Leiden for making me feel so good here.



# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
<b>2</b>	<b>Methodology</b>	<b>15</b>
2.1	Time-dependent confounding . . . . .	15
2.2	Marginal Structural Models (MSMs) . . . . .	17
2.2.1	Estimation method . . . . .	19
2.2.2	Assumptions . . . . .	26
2.3	Applications of MSM . . . . .	27
2.4	Alternative methods for causal inference . . . . .	29
<b>3</b>	<b>Data description</b>	<b>31</b>
3.1	Data gathering process . . . . .	31
3.2	Descriptive statistics . . . . .	35
3.3	Exposures of interest . . . . .	40
3.4	Exposure-confounder feedback . . . . .	45
3.5	Missing values, data preprocessing, patients' eligibility . . . . .	50
3.5.1	Missing values . . . . .	51
3.5.2	Data preprocessing . . . . .	51
3.5.3	Patients' eligibility for analysis . . . . .	52
<b>4</b>	<b>MSMs for Histological Response</b>	<b>55</b>
4.1	MSM for causal effect of cycle delays on histological response – a simplified example . . . . .	55
4.1.1	Investigation of time-dependent confounding and exposure – confounder feedback . . . . .	57
4.1.2	Model formulation and estimation . . . . .	60
4.1.3	Results and discussion . . . . .	62
4.2	MSM for joint causal effect of cycle delays and dose reductions on histological response . . . . .	63
4.2.1	Investigation of time-dependent confounding and exposure – confounder feedback . . . . .	66
4.2.2	Model formulation and estimation . . . . .	73
4.2.3	Results and discussion . . . . .	76

4.3	MSM for joint causal effect of cycle and surgery delays, and dose reductions on histological response . . . . .	79
4.3.1	Investigation of surgery delay confounding . . . . .	80
4.3.2	Model formulation and estimation . . . . .	80
4.3.3	Results and discussion . . . . .	87
<b>5</b>	<b>MSCPHMs for survival</b>	<b>91</b>
5.1	Investigation of time-dependent confounding and exposure – confounder feedback . . . . .	92
5.2	Model formulation and estimation . . . . .	98
5.3	Results and discussion . . . . .	106
<b>6</b>	<b>Discussion</b>	<b>109</b>
6.1	Overview of the results and their limitations . . . . .	109
6.2	Study limitations due to the available data . . . . .	111
6.2.1	Limitations due to the confounders . . . . .	111
6.2.2	Limitations due to the exposures . . . . .	113
6.3	Data hierarchy . . . . .	114
6.4	Simulation study . . . . .	114
<b>A</b>	<b>R code</b>	<b>121</b>
A.1	R code for Chapter 3 . . . . .	121
A.2	R code for Chapter 4 . . . . .	141
A.3	R code for Chapter 5 . . . . .	160

# List of Figures

2.1	Causal graphs for time-dependent exposure . . . . .	16
2.2	DAG describing the causal relationships presented in Example 2.2 . . . . .	18
3.1	Control arm treatment and monitoring schedule . . . . .	32
3.2	Chemotherapy form . . . . .	34
3.3	Plots of key demographic variables . . . . .	35
3.4	Plots of protocol treatment completion . . . . .	36
3.5	Plots of surgery timing in cycles and in days since completion of the last preoperative cycle . . . . .	37
3.6	Plots of histological response, overall and event-free survival . . . . .	38
3.7	Barplots of Common Terminology Criteria for Adverse Events (CTCAE) grades per toxicity and cycle . . . . .	39
3.8	Plots of different aspects of treatment duration . . . . .	41
3.9	Plots of delays and associated reasons for delay . . . . .	43
3.10	Scatterplots of standardised dose of CDDP vs. standardised dose of DOX per cycle . . . . .	44
3.11	Relationship between exposure, confounder and outcome . . . . .	46
3.12	Therapy modifications per cycle as severest toxicity from previous cycle . . . . .	48
3.13	Therapy modifications per cycle as number of toxicities from previous cycle with grade larger than one . . . . .	49
3.14	Treatment trajectories of patients 11 and 1106, and therapy of patient 1106 according to the protocol . . . . .	50
3.15	Plots of examples of records that should be treated with care . . . . .	52
3.16	Flow-chart of patients selected for analyses . . . . .	53
4.1	DAG describing the causal relationships between cycle delays, confounders and histological response . . . . .	56
4.2	DAG describing the causal relationships between exposures, confounders and histological response . . . . .	65
4.3	Temporal order within a cycle and interrelationships between exposures and confounders . . . . .	66
4.4	Boxplots of subject-specific weights and its components . . . . .	77
4.5	DAG describing the causal relationships between exposures, confounders and histological response . . . . .	79

4.6	Boxplots of the final weights and all its components . . . . .	86
5.1	DAG describing the causal relationships between exposures, confounders and survival . . . . .	92
5.2	Boxplots of the final weights and all its components . . . . .	105

# List of Tables

4.1	Toxicities effect estimates on histological response . . . . .	59
4.2	Toxicities effect estimates on allocation of cycle delays . . . . .	59
4.3	Cycle-two delay effect estimates on occurrence of severe toxicities . . . . .	60
4.4	Distribution of estimated probabilities of cycle-two delay, probabilities of observed exposure, and final weights . . . . .	62
4.5	Causal and association odds-ratio estimates for the effect of cycle-two delay on histological response . . . . .	63
4.6	Distribution of number of preoperative cycles . . . . .	63
4.7	Distribution of cumulative exposure . . . . .	68
4.8	Toxicities effect estimates on histological response . . . . .	69
4.9	Toxicities effect estimates on administration of delays . . . . .	69
4.10	Toxicities effect estimates on administration of dose reductions . . . . .	71
4.11	Effect estimates of exposure on subsequent occurrence of toxicities . . . . .	72
4.12	Distribution of estimated probabilities of observed exposure and weights . . . . .	77
4.13	Causal and association odds-ratio estimates for the effects of cycle delay, dose reductions, and length of preoperative treatment on histological response . . . . .	78
4.14	Effect of last preoperative cycle toxicities on surgery delay . . . . .	80
4.15	Models for allocation of cycle delays . . . . .	82
4.16	Models for allocation of dose reductions . . . . .	82
4.17	Distribution of recorded and missing HRe by country of treatment . . . . .	87
4.18	Causal and association odds-ratio estimates for the joint effect of cycle delays, dose reductions, surgery delay, and the number of preoperative cycles on histological response . . . . .	88
5.1	Toxicities effect estimates on EFS . . . . .	94
5.2	Toxicities effect estimates on administration of delays . . . . .	95
5.3	Toxicities effect estimates on administration of dose reductions . . . . .	96
5.4	Toxicities effect estimates on allocation of surgery delay . . . . .	97
5.5	Effect estimates of exposure on subsequent occurrence of toxicities . . . . .	98
5.6	Models for allocation of cycle delays using all treatment data . . . . .	102
5.7	Models for allocation of dose reduction using all treatment data . . . . .	102
5.8	Causal and association hazard-ratio estimates for the joint effect of cycle delays, dose reductions, and surgery delay on event-free survival . . . . .	106

5.9 Causal and association hazard-ratio estimates for the joint effect of cycle delays, dose reductions, and surgery delay on overall survival . . . . . 107

# List of Abbreviations

<b>AIC</b>	Akaike's information criterion
<b>CDDP</b>	CISPLATIN
<b>CTCAE</b>	Common Terminology Criteria for Adverse Events
<b>DAG</b>	Directed Acyclic Graph
<b>DOX</b>	DOXORUBICIN
<b>EFS</b>	Event-Free-Survival
<b>GR</b>	Good (histological) Response
<b>G-CSF</b>	Granulocyte Colony-Stimulating Factor
<b>HRe</b>	Histological response
<b>IPCW</b>	Inverse Probability-Of-Censoring Weight
<b>IPTW</b>	Inverse Probability-Of-Treatment Weight
<b>MSCPHM</b>	Marginal Structural Cox Proportional Hazards Model
<b>MSM</b>	Marginal Structural Model
<b>OS</b>	Overall Survival
<b>PR</b>	Poor (histological) Response
<b>RCS</b>	Restricted cubic spline
<b>SD</b>	Standard deviation
<b>SE</b>	Standard error
<b>SNM</b>	Structural Nested Model
<b>WBC</b>	White blood cell



# Chapter 1

## Introduction

The most widely used statistical analysis technique, regression analysis, quantifies the association between an outcome variable and an exposure. Due to the complexity of the real-world processes, this relationship is often confounded, i.e. there exists a third party which influences both the outcome and the exposure. The effect of the confounder can be corrected for by including the confounder just as another covariate into the model. In this way, we no longer assess the strength of the overall relationship between the exposure and the outcome. Instead, we purify this association by correcting for the influence of the confounder on both the exposure and the outcome, and thus claim to assess the direct causal relationship between the exposure and the outcome. However, the confounder is often unknown and/or unmeasured. To overcome this problem, scientists began to perform controlled experiments. In the heart of an experiment lays the concept of randomisation. Through random allocation of exposure to experimental units the effect of the confounder is neutralised, because the exposure is assigned independently of the confounder. This setting allows one to interpret the association parameters in a causal way.

Nowadays data from randomised experiments is used to assess relationships beyond the primary goal of the study. From the perspective of secondary analysis, the exposure of interest is no longer allocated at random to the experimental units, and the existence of confounders is almost certain. Methods have been developed such that even in such cases we can make causal inference.

Studies in clinical research usually have longitudinal nature. In this case the confounding may be time-dependent. That is, the value of the confounder (as well as the exposure) may change over time.

In this thesis we use the data from a randomised controlled trial for chemotherapy in osteosarcoma to illustrate the causal inference methodology that needs to be used in the presence of time-dependent confounding. In addition, we show that the exposure influences future values of the confounders. This phenomenon is referred to as *exposure-confounder feedback*.

The aim of this thesis is to adapt the methodology for causal inference in presence of time-dependent confounding and exposure-confounder feedback to the complex set-up of

data from a randomised controlled trial in chemotherapy for osteosarcoma. We discuss the methodology through analyses of binary and survival outcomes.

The thesis is structured as follows. In Chapter 2 we present the relevant methodology, and a literature review. An extensive description of the data is given in Chapter 3. For illustration purposes Chapter 4 starts with building a simple model for binary exposure, which is then extended through several steps to finally address all relevant research questions. Chapter 5 guides the reader through the process of building models for time-to-event outcomes. Finally, in Chapter 6 we present some conclusions based on the performed analyses, their limitations, and suggestions for future research.

The statistical analyses are performed in the R-software environment. All R code can be found in the appendix.

## Chapter 2

# Methodology

This chapter reviews the methodology for causal inference with Marginal Structural Models (MSMs). It starts with an introduction of the type of problems, which require the use of MSMs. Section 2.2 is meant to guide the reader through the process of building and estimating a MSM. As such, it is a systematic review of all, to the extent of writer's knowledge, published caveats concerning MSMs. The section ends with a list of assumptions that guarantee the validity of a MSM. Further, a number of examples of the use of MSMs in practice are discussed in Section 2.3. Finally, Section 2.4 is devoted to two alternative methods for causal inference, and to contrast them with MSMs.

### 2.1 Time-dependent confounding

In order to address the problem of time-dependent confounding it is useful to consider an example.

**Example 2.1.** Let  $Y$  be a dichotomous outcome variable taking value 1 if the percentage of viable tumour in a resected specimen is less or equal to 10%, called *good histological response* (GR), and 0 otherwise, referred to as *poor histological response* (PR). Let  $A_k$  indicate the occurrence of a dose reduction (i.e.  $A_k$  is a binary variable) during the  $k$ -th preoperative chemotherapy cycle, for  $k = 1, 2$ . The goal here is to estimate the causal effect of the time-dependent exposure  $A_k$  on the outcome  $Y$ .

Suppose the presence of myelosuppression denoted by  $L$  is also recorded. Then for  $k = 1, 2$ ,  $L_k$  takes values 0 or 1 in case of absence or occurrence of myelosuppression, respectively. Finally, let  $U_k$  represent the collection of all possible treatment side effects which were not measured.

Figure 2.1 illustrates three different scenarios involving  $Y$ ,  $A_k$ ,  $L_k$ , and  $U_k$ . Each sub-figure represents causal relationships between the variables through a Directed Acyclic Graph (DAG), where the variables are located at the vertices (nodes), and the directed edges (arrows) visualise direct causal effects.

**Definition 2.1. Confounder** A covariate  $L$  is a confounder of the effect of exposure  $A$  on an outcome  $Y$  if it predicts both  $A$  and  $Y$ .

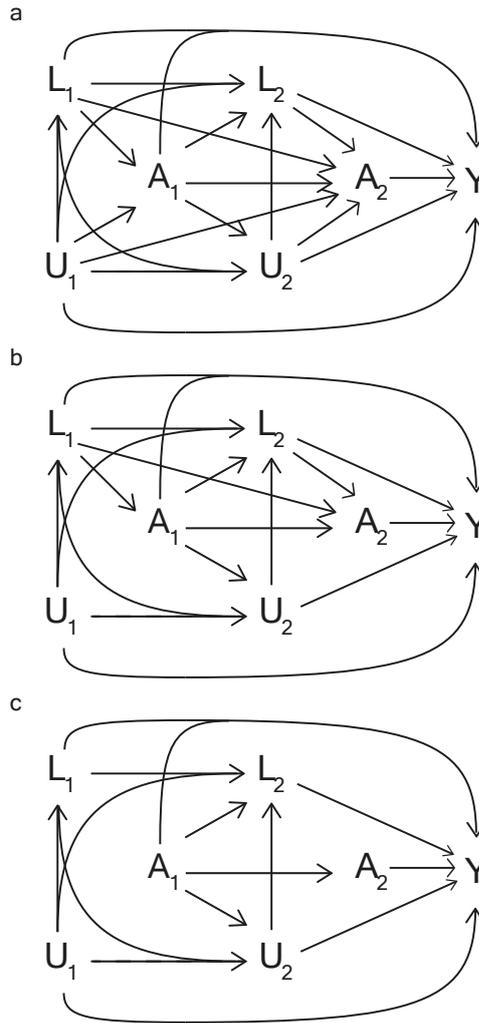


Figure 2.1: Causal graphs for time-dependent exposure

**Definition 2.2. Time-dependent confounder** An independent risk factor is a time-dependent confounder of the effect of exposure on an outcome if (1) it predicts subsequent exposure, and (2) past exposure predicts subsequent risk factor level.

According to Definition 2.2,  $L_k$  is a time-dependent confounder of the effect of  $A_k$  on  $Y$  in Figures 2.1a and 2.1b. Figure 2.1a visualises a causal graphs of time-dependent measured and unmeasured confounding. On the other had, there is no unmeasured confounding given the data on measured confounders  $L_k$  if, as in Figure 2.1b, no edges connect  $U_k$  with  $A_k$ . Finally, Figure 2.1c, lacks all edges between any risk factor,  $L$  or  $U$ , and any  $A$ -vertice, and thus represents the case of no confounding. In such a situation, estimated association parameters can be interpreted in a causal way.

In observational studies the researcher cannot ascertain the lack of unmeasured con-

founding from the available data. However, the data can be used to empirically test whether the effect of the exposure on the outcome is confounded by the measured covariates.

More examples of time-dependent confounding are discussed in Section 2.3.

## 2.2 Marginal Structural Models (MSMs)

This section presents the methodology concerning MSMs. For this purpose, two main examples are used. Estimation methods for MSMs comprise the body of the section. In addition, we address some caveats associated with the formulation and estimation of MSMs. At the end of this section, we review the assumptions of MSMs.

MSMs, introduced by Robins [1] in his seminal paper, are convenient methods for causal inference in the presence of *time-dependent confounding*. They closely resemble traditional regression models used to quantify associations. The major difference, though, is that MSMs are used to model the expected value of *counter-factual outcomes*, rather than *factual, observed outcomes*.

As in other mathematical literature, throughout this thesis random variables are denoted by capital letters, e.g.  $Y, A, L$ , and their realisations with lower-case letters, e.g.  $y, a, l$ . Further, **boldface**, e.g.  $\mathbf{V}$ , is used to distinguish vectors from univariate variables.

**Definition 2.3. Treatment history** A hypothetical treatment history  $\bar{a}$  is a pattern of time-dependent exposure values from the beginning of a treatment until its end.

That is, in the context of Example 2.1, the treatment pattern  $(A_1 = a_1, A_2 = a_2)$  forms the treatment history  $\bar{a} = (a_1, a_2)$ .

**Definition 2.4. Counter-factual outcome** A counter-factual outcome  $Y_{\bar{a}}$  is the outcome that a patient would have experienced if he/she has been, possibly contrary-to-fact, exposed to a treatment history  $\bar{a}$ .

The number of counter-factual outcomes depends on the set of possible combinations of exposure values throughout the studied period. In Example 2.1, the exposure could take two possible values at each of the two time points. This amounts to  $2^2 = 4$  potential treatment trajectories, each of which corresponds to a different counter-factual outcome. In practice each patient follows only one of the possible treatment trajectories, and his/her outcome is observed only once. We denote observed outcome by  $Y$ . This brings us to the fundamental assumption of MSMs, which links factual to counter-factual outcomes.

**Consistency** Let  $Y$  be the outcome of a given patient allowing treatment trajectory  $\bar{A}$ . Consistency assumes  $Y = Y_{\bar{a}}$  if  $\bar{A} = \bar{a}$ , where  $Y$  is the random variable corresponding to the factual outcome,  $Y_{\bar{a}}$  denotes the counter-factual outcome of treatment history  $\bar{a}$ . Then all counter-factual outcomes for which  $\bar{A} \neq \bar{a}$  are unobserved.

A Marginal Structural Logistic Model for the causal effect of treatment history on the expected value of a binary outcome, as in Example 2.1, is defined as follows:

$$E[Y_{\bar{a}}] = g(\bar{a}, \boldsymbol{\beta}),$$

where

$$g(\bar{a}, \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1 f(\bar{A}))}{1 + \exp(\beta_0 + \beta_1 f(\bar{A}))}, \quad (2.1)$$

$E[\cdot]$  denotes mathematical expectation,  $\boldsymbol{\beta} = (\beta_0, \beta_1)$ , and  $f(\bar{A})$  is a function of treatment history  $\bar{A}$  (usually a form of cumulative exposure). Model (2.1) is called *marginal* because it models the marginal distribution of counter-factual outcomes, rather than their joint distribution. Further, MSMs are referred to as *structural* models because causal models in econometric and social sciences are often called structural [2].

Parameter  $\beta_1$  in (2.1) quantifies the average causal effect of treatment history  $\bar{A}$  on the outcome. That is, the difference in counter-factual outcomes under regimens  $\bar{a}_1$  and  $\bar{a}_2$  is a function of  $\beta$ ,  $\bar{a}_1$ , and  $\bar{a}_2$ . In the context of Example 2.1, if we model treatment history as  $f(\bar{A}) = \sum_{k=1}^2 A_k$ , then  $f(\bar{A})$  takes values 0, 1, or 2. The corresponding  $\beta_1$ -parameter estimates the causal odds ratio of a Good (histological) Response (GR) for an additional occurrence of dose reduction. Different model parametrisation will, of course, lead to different interpretation of the parameter estimates.

MSM for other types of outcomes could also be defined. Let us consider an extension of Example 2.1 where the outcome is time-to-event.

**Example 2.2.** We are interested in the causal effect of dose reductions during chemotherapy on survival of patients after completion of the therapy. Let  $T$  be patient's time to death in months since completion of six chemotherapy courses equally-spaced in time. Here the treatment history,  $f(\bar{A}) = f(A_1, A_2, A_3, A_4, A_5, A_6)$ , could be expressed as the number of courses at which the dose was reduced, i.e.  $f(\bar{A}) = \sum_{k=1}^6 A_k$ . If we condition survival on completion of the therapy, the treatment history is baseline, i.e. the time-counting process begins after we have observed the exposure. Additionally, before administering the next chemotherapy cycle, clinicians determine the occurrence of myelosuppression. That is, for each course of chemotherapy  $k$   $L_k$  takes value 1 if myelosuppression is present, and 0 otherwise ( $k = 1, 2, \dots, 6$ ). If myelosuppression is present, usually the dose of chemotherapeutic agents to be given is decreased. At the same time myelosuppression is a prognostic factor for survival. This implies that the effect of dose reductions on survival is confounded by myelosuppression.

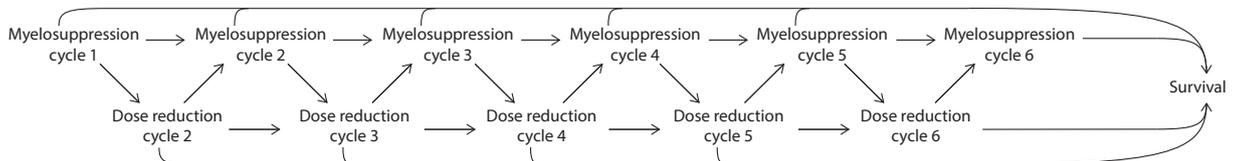


Figure 2.2: DAG describing the causal relationships presented in Example 2.2

To further clarify Example 2.2 the corresponding DAG is presented in Figure 2.2. *Myelosuppression in cycle  $k$*  predicts *dose reduction in cycle  $k + 1$* , *myelosuppression in cycle  $k + 1$* , and patient's *survival*. However, *dose reduction in cycle  $k$*  predicts occurrence of *myelosuppression in cycle  $k + 1$* , *dose reduction in cycle  $k + 1$*  and the *survival*. Each of these relationships is visualised with an edge between dependent vertices.

A Marginal Structural Cox Proportional Hazards Model (MSCPHM) defined as:

$$\lambda_{T_{\bar{a}}}(t) = \lambda_0(t) \exp(\beta f(\bar{A})), \quad (2.2)$$

is employed in this setting. The term  $T_{\bar{a}}$  denotes subject's time to event had he/she followed a treatment history  $\bar{a}$ ,  $\lambda_{T_{\bar{a}}}(t)$  is the hazard of  $T_{\bar{a}}$  at month  $t$  since therapy completion,  $\lambda_0(t)$  is a time-dependent baseline hazard, and  $\exp(\beta)$  is the causal rate ratio for the effect of treatment on survival. In other words,  $\exp(\beta)$  quantifies how more frequently patients die, on average, if the number of therapy courses with a reduced dose is increased by one.

### 2.2.1 Estimation method

Models (2.1) and (2.2) do not contain any functional form of the time-dependent confounder,  $L$ , or other covariates because they are models for causal effects on the entire study population. The  $\beta$ -parameters of these models could be consistently estimated from weighted regression models for factual outcomes. Prior to discussing why this is the case, it is useful to consider the notions of *causal* and *statistical exogeneity*.

**Definition 2.5. Causal exogeneity** A treatment process is *causally exogenous* if the conditional probability of receiving treatment  $A_k$  for time interval  $k$  given past treatment and prognostic factor history (measured and unmeasured) depends only on past treatment history  $\bar{A}_{k-1}$  [2]; that is,

$$P(A_k \mid \bar{A}_{k-1}, \bar{L}_k, \bar{U}_k) = P(A_k \mid \bar{A}_{k-1}).$$

Practically, causal exogeneity translates to lack of measured and unmeasured confounding, which turns the association parameters into causal ones. However, causal exogeneity cannot be assessed, while its complement, statistical exogeneity, can.

**Definition 2.6. Statistical exogeneity** A treatment  $A_k$  is a *statistically exogenous* process if the probability of receiving treatment for time interval  $k$  does not depend on the history of measured time-dependent prognostic factors up to time interval  $k$  conditional on treatment history before  $k$  [2]; that is,

$$\bar{L}_k \perp\!\!\!\perp A_k \mid \bar{A}_{k-1}.$$

*Causal exogeneity* implies *statistical exogeneity*, but the reverse is not true since statistical exogeneity refers only to lack of measured confounding, while causal exogeneity excludes both measured and unmeasured confounding. In other words, a process might be statistically exogenous but not causally exogenous if there are unmeasured prognostic

factors that predict the probability of a treatment  $A_k$  for time interval  $k$  given past treatment history [3].

One could quantify the degree to which a treatment process is statistically nonexogenous through time interval  $j$  by computing the random quantity,

$$SW_j = \prod_{k=1}^j \frac{f[A_k|\bar{A}_{k-1}]}{f[A_k|\bar{A}_{k-1}, \bar{L}_k]}, \quad (2.3)$$

for  $j$  in  $1, 2, \dots, 6$ , where  $f[A_k|\bar{A}_{k-1}, \bar{L}_k]$  is the conditional probability density or mass function of a continuous or discrete exposure, respectively, evaluated at observed exposure and confounder history of each subject. The quantity  $SW_j$  is referred to as *stabilised inverse-probability-of-treatment weights* (IPTW), or *stabilised weights* [2]. The denominator in Equation (2.3) is the probability for a specific patient to receive the observed treatment,  $A_k$ , for time interval  $k$  given his/her treatment and covariate history. In terms of Example 2.2, it is the probability of administration of dose reduction through chemotherapy cycle  $k$  given dose reductions and occurrences of myelosuppression from cycles 1 to  $k - 1$  and from cycle 1 to  $k$ , respectively. The numerator in Equation (2.3) is the probability for a specific patient to receive the observed treatment,  $A_k$ , for time interval  $k$  given only his/her treatment history.

The rationale behind the use of inverse-probability-of-treatment weights in the context of time-dependent confounding is as follows. A *pseudo-population*, where each subject from the original sample contributes  $SW_j$  copies of him-/herself for time interval  $j$ , is a pseudo-replication of the original sample in which there is no time-dependent measured confounding. Moreover, the causal effect estimates from the pseudo-population equal those in the actual population [4]. A simple example in the context of Example 2.1 could be as follows: suppose that occurrence of myelosuppression in the first preoperative chemotherapy cycle is associated with allocation of dose reduction in the second cycle. That is, (1) the sample contains more examples of both, myelosuppression ( $L_1 = 1$ ) and dose reduction ( $A_2 = 1$ ), than dose reduction without myelosuppression ( $L_1 = 0$ ) and/or more cases without both myelosuppression and dose reduction than lack of dose reduction after a course of myelosuppression. Then the effect of dose reduction on the outcome, Histological response (HRe), is confounded by myelosuppression. The estimated probability of dose reduction in the second preoperative cycle given previous exposure and myelosuppression history, i.e.  $P(A_2 = 1|A_1 = a_1, L_1 = l_1)$ , will be high if  $L_1 = 1$ , and low if  $L_1 = 0$ . Then IPTW will be small (large) for high (low) probability of exposure since the two are inversely proportional. That is, (2) in the pseudo-population cases without myelosuppression but with dose reduction will be over-represented. Conditions (1) and (2) imply that in the pseudo-population the distribution of myelosuppression is equal among patient with reduced and full dose, i.e. myelosuppression is no longer associated with allocation of exposure. Finally, the latter translates into lack of confounding in the pseudo-population.

Furthermore, the weights are subject-specific since they are a function of the subject's covariates. It follows that the exposure appears to be randomly assigned across

patients with similar characteristics. In this way the pseudo-population created by Inverse Probability-Of-Treatment Weights (IPTWs) mimics a randomised clinical trial. As a result the association parameter estimates have causal interpretation.

The example above addresses a situation where patients are weighted with the inverse probability of observed treatment, i.e.  $\text{weight} = 1/P(A_1 = a_1|A_0 = a_0, L_0 = l_0)$ . Therefore, observations that represent counter-examples of the observed association between exposure and confounder may get very large weights. The estimated inverse probability-of-treatment is dependent on the sample at hand, and it is rather unstable especially when there is a strong association between exposure and confounder [4]. Instead, stabilised weights as (2.3) can then be used. Contrary to unstabilised weights, the numerator of stabilised weights considers the probability of observed exposure given exposure history only. This modification aims to decrease the difference between numerator and denominator, and therefore limits the variance of the estimated weights. The difference between the two components of the weights reflects only the confounding due to time-varying confounders [5]. Further stabilisation of the weights could be achieved through truncation. This possibility is discussed in Section 2.2.2 as it relates to *Positivity* assumption.

The quantity  $SW_j$  in (2.3) is equal to 1 for all time intervals  $j$  if and only if the treatment process  $A$  is *statistically exogenous* [2]. Another way to investigate whether time-dependent measured confounding is present, is to fit three regression models for association. These models study whether edges between  $L_k$  and  $Y$ ,  $L_k$  and  $A_{k+1}$ , and  $A_k$  and  $L_{k+1}$  are present. In the context of Example 2.1, to determine whether the occurrence of myelosuppression is an independent predictor of HRe one could fit a logistic regression model with  $L_1$  and  $L_2$  as the only predictors of HRe. If the effects of the latter are supposed to be confounded with some baseline patient characteristics, those should also be included into the model. A significant association parameter estimate confirms the hypothesis that the potential confounder is a predictor of the outcome. In case there are more than one potential confounders, a multiple logistic regression model might be estimated to evaluate the impact of each possible predictor corrected for the others. To identify whether current value of myelosuppression predicts subsequent administration of dose reduction one could estimate a pooled logistic regression model. In this model the exposure is regressed on the potential confounder, while each preoperative patient-cycle is treated as an independent observation.

Correct estimation of the weights is the cornerstone of MSMs. Therefore it is important to model the temporal order of occurrence of exposure and confounder, as well as their interdependence in a proper way. Two simple examples follow.

In studies of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on survival of HIV-infected men [2], the value of the time dependent confounder, CD4 count, is measured at each patient visit  $t$ , for  $t$  taking discrete values  $1, 2, \dots, 16$ . The value of the confounder at visit  $t$  determines the administration of prophylaxis therapy for period from visit  $t$  to  $t + 1$ . That is,  $L_t$  predicts  $A_t$ , and the confounder at time  $t$  is observed before the exposure at time  $t$ . In chemotherapy studies [6], toxicities throughout cycle  $k$  are time-dependent confounders for the effect of therapy modifications on histological

response. They occur as a result of chemotherapy cycle  $k$ , and predict the exposure in cycle  $k + 1$  (with  $k = 1, 2$ ). Note that in chemotherapy studies there is a one cycle lag between the moments at which the dependent exposure and confounder are observed. These details are crucial for the validity of the MSM.

Equation (2.3) is not the only possible formulation of stabilised weights. The way the weights are computed depends on the research question at hand. By definition,  $L_1$  is a vector which contains baseline covariates like patients' age, gender, and other relevant factors. For the sake of generality, vector notation, i.e. boldface  $\mathbf{L}_1$ , is suppressed in most of the formulas, and, where applied, is used to stress the nature of the elements. If the researcher is interested in, for example, gender as effect modifier, the MSM and the models used to estimate the IPTWs should be updated accordingly. That is, the MSM should include gender main effect and interaction term between gender and the exposure. The model used to estimate the probability of exposure given exposure history, the numerator in (2.3), should be augmented to include all potential effect modifiers [2]. The formula for stabilised weights then becomes

$$SW_j = \prod_{k=0}^j \frac{f[A_k | \bar{A}_{k-1}, \mathbf{V}]}{f[A_k | \bar{A}_{k-1}, \bar{\mathbf{L}}_k]}, \quad (2.4)$$

where  $\mathbf{V}$  is used to denote a vector of all potential baseline effect modifiers, and  $\mathbf{V} \in \mathbf{L}_1$ .

MSMs cannot be used to estimate interaction effects of treatment with time-dependent covariate. If this is the aim of the analysis, Structural Nested Model (SNM), discussed in Section 2.4, should be used.

Weights in Equation (2.4) are more stable than weights in Equation (2.3) but caution is required when they are used. Weights computed according to (2.4) create a pseudo-population in which, for each time interval  $j$ , exposure is randomised only within the levels of each of the covariates in  $\mathbf{V}$  [5]. For more insight into why this is the case, let us re-parametrise Equation (2.4). Let  $\mathbf{Z}_k = (A_k, \mathbf{V})$ . Then weights computed as  $P(\mathbf{Z}_k | \bar{\mathbf{Z}}_{k-1})/P(\mathbf{Z}_k | \bar{\mathbf{Z}}_{k-1}, \bar{\mathbf{L}}_k)$  create a pseudo-population in which  $\bar{\mathbf{L}}_k \perp\!\!\!\perp \mathbf{Z}_k | \bar{\mathbf{Z}}_{k-1}$ . For simplicity assume that  $\mathbf{Z}_k$  is a vector of two binary variables: gender ( $G$ ) and dose reduction in time interval  $k$  ( $A_k$ ). It follows that  $\bar{\mathbf{L}}_k \perp\!\!\!\perp (G, A_k) | (G, \bar{A}_{k-1})$ . That is,  $\bar{\mathbf{L}}_k$  is independent of  $A_k$  given  $\bar{A}_{k-1}$  within the same level of  $G$  but not across the levels of  $G$ , because it is conditional on  $G$ . Therefore, in the pseudo-population the effect of the exposure could still be confounded with baseline covariates in  $\mathbf{V}$  (in the example gender). Accordingly, when using weights as in (2.4), one should always include all baseline covariates in  $\mathbf{V}$  in the MSM. In the literature,  $\mathbf{V}$  usually contains baseline effect modifiers. In such cases, interaction terms between  $\mathbf{V}$  and the exposure are additionally included in the MSM. As a result, such MSMs are no longer models for marginal, i.e. unconditional, causal effects on the whole study population.

The Marginal Structural Cox Proportional Hazards Model (MSCPHM) (2.2) disregards censoring, which is a common phenomenon in survival analysis. To adjust for right censoring due to loss to follow-up is only to say that we are interested in estimating the causal effect of treatment history  $\bar{a}$  on survival in absence of censoring. Let  $C_k$  be a binary variable taking value 0 in study visit  $k$  if a patient is in the risk set at time  $k$ ,

and 1 otherwise. In this formulation  $k$  takes integer values and denotes the number of equally-spaced in time patient-visits to the hospital (as in the studies of HIV-positive men [2, 7]). In these trials follow-up starts from visit 1, and treatment is updated every visit based on time-dependent confounder's value, available at the beginning of each visit  $k$ . Censoring is usually assumed to occur in the end of period  $k$ , i.e.  $C_k$  translates into censoring just before  $k + 1$  starts. We say that censoring is ignorable or non-informative if the conditional hazard of being censored at visit  $k$  does not depend on the death times  $T_{\bar{a}, \bar{c} \equiv 0}$  given treatment and confounder histories [2]. Under the assumption of ignorable censoring and no unmeasured confounding, causal effects of time-dependent treatments on survival could be estimated through weighted Cox regression on factual outcome. Patient's contribution to the risk set at time  $t$  is weighted by  $SW(t) \times SW^\dagger(t)$ , given as

$$SW(t) = \prod_{k=0}^t \frac{f[A_k | \bar{A}_{k-1}, \bar{C}_k = 0, T \geq k]}{f[A_k | \bar{A}_{k-1}, \bar{L}_k, \bar{C}_k = 0, T \geq k]}, \quad (2.5)$$

and

$$SW^\dagger(t) = \prod_{k=0}^t \frac{P[C_k = 0 | \bar{C}_{k-1} = 0, \bar{A}_{k-1}, T > k]}{P[C_k = 0 | \bar{C}_{k-1} = 0, \bar{A}_{k-1}, \bar{L}_k, T > k]}. \quad (2.6)$$

These are subject-specific time-dependent weights, where patient index is suppressed for simplicity. Equation (2.5) defines the Inverse Probability-Of-Treatment Weight (IPTW), while (2.6) Inverse Probability-Of-Censoring Weight (IPCW). Since IPTW is conditioned on not being censored,  $f[A_k | \bar{C}_k = 0, \dots]$ , the product of the two probabilities is a decomposition of their joint probability – the probability of remaining uncensored and experiencing exposure  $A_k$  for each time period  $k$  from 0 to  $t$ .

In Equations (2.5) and (2.6) conditions  $T \geq k$  and  $T > k$  respectively indicate that in the pseudo-population treatment trajectories are as long as in the original sample because weights are computed until the actual event or censoring time  $T$ .

As discussed before, causal effects could be consistently estimated by weighted regression models for factual outcomes on pseudo-population created by IPTWs. The weights estimated through (2.5) and (2.6) depend on time. When dealing with time-to-event data, pooled logistic regression is often used in the literature since few statistical software packages can fit a Cox proportional hazards model with weights that depend on time. This methodology treats each patient-visit as an independent observation. The equivalence of Cox proportional hazards and pooled logistic regression models was extensively discussed in [8, 9]. The following model is then estimated:

$$P(D(k) = 1 | \bar{D}(k-1) = 0, \bar{A}(k-1)) = \frac{\exp(\gamma_0 + \gamma_1 f(\bar{A}))}{1 + \exp(\gamma_0 + \gamma_1 f(\bar{A}))}, \quad (2.7)$$

where  $D(k)$  is an indicator variable equal to 1 if the patient died between visit  $k - 1$  and visit  $k$ , and 0 when the patient was alive. When Model (2.7) is fit on a pseudo-population created by weights  $SW(t) \times SW^\dagger(t)$  parameters  $\gamma_0$  and  $\gamma_1$  are consistent estimates of the  $\beta$ -parameters of the corresponding MSM:

$$P(D_{\bar{a}, \bar{c}=0}(k) = 1 | \bar{D}_{\bar{a}, \bar{c}=0}(k-1) = 0) = \frac{\exp(\beta_0 + \beta_1 f(\bar{A}))}{1 + \exp(\beta_0 + \beta_1 f(\bar{A}))}, \quad (2.8)$$

In the MSM (2.8)  $D_{\bar{a}, \bar{c}=0}(k)$  denotes the counter-factual outcome {dead/alive} at visit  $k$  had the patient, possibly *contrary-to-fact*, followed treatment history  $\bar{a}$ , and remained alive and uncensored up to visit  $k$ . Note the change in notation of time component  $k$  which only applies to Equations (2.7) and (2.8).

Following the same argument, we fit model

$$P(Y = 1 | \bar{A} = \bar{a}) = \frac{\exp(\gamma_0 + \gamma_1 f(\bar{A}))}{1 + \exp(\gamma_0 + \gamma_1 f(\bar{A}))} \quad (2.9)$$

to a pseudo-population created by weights as in (2.3) to consistently estimate the parameters of the marginal structural logistic Model (2.1).

However, the regression procedures that estimate Models (2.7) and (2.9) do not take into account the fact that the weights were also estimated. This invalidates the standard errors of the parameter estimates and the confidence intervals fail to withhold the desired coverage [4, 2, 7]. To overcome this problem robust standard errors have to be estimated [1, 3]. This could be achieved by using the so-called “sandwich” covariance matrix estimator, or fitting the model within the framework of generalised estimating equations and specifying independence working covariance matrix.

Estimation of stabilised weights  $SW_j$  for interval  $j$  is a challenging task. First the probabilities that form the weights need to be estimated. The way they are modelled depends on the functional form of exposure history,  $f(\bar{A})$ , in the MSM. The choice of the latter should be driven by the prespecified research question that motivates the analysis, taking into account interpretability of the MSM. In order to emphasise the importance of this aspect and its dependence on the research setting, throughout this chapter the general form of  $f(\bar{A})$  is used. Often one of the following two alternative formulations of  $f(\bar{A})$  are used. One is to unfold exposure and assess its effect in every instance of measurement separately, e.g.  $\beta_1 f(\bar{A}) = \beta_{1,1} A_1 + \beta_{1,2} A_2$ , and the other is to use some form of cumulative exposure, e.g.  $\beta_1 f(\bar{A}) = \beta_1 f(A_1, A_2) = \beta_1 \sum_{j=1}^2 A_j$ .

Within the context of Example 2.1, one might estimate the causal effect of an additional administration of dose reduction on HRe. This could be achieved by modelling cumulative exposure as the sum of dose reduced cycles, i.e.  $f(\bar{A}) = \sum_{k=0}^1 A_k$  and takes values 0, 1, or 2 for two preoperative cycles. The probability of the observed exposure in the denominator of (2.3) can be estimated by fitting a pooled logistic regression model

$$p_k = P(A_k = 1 | A_{k-1}, L_{k-1}) = \frac{\exp(\alpha_{0,k} + \alpha_1 A_{k-1} + \alpha_2 L_{k-1})}{1 + \exp(\alpha_{0,k} + \alpha_1 A_{k-1} + \alpha_2 L_{k-1})}, \quad (2.10)$$

where  $\alpha_{0,k}$  is a cycle-specific intercept, and  $A_{k-1}$  and  $L_{k-1}$  are indicator variables of dose reduction and myelosuppression in cycle  $k-1$ , respectively. In this example there are only two periods of time, cycle 1 and 2, and the cycle-specific intercept could be

modelled with an overall intercept,  $\alpha_{0,0}$ , and an indicator variable for the second cycle, i.e.  $\alpha_{0,1} = \alpha_{0,0} + \alpha'_{0,1} \mathbb{1}_{\{\text{second cycle}\}}$ .

In research settings with more time points/intervals, either more indicator variables should be created, or a function of time could be used to define time-specific intercepts. As mentioned earlier, whether  $L_{k-1}$  or  $L_k$  is used to model the exposure depends on the temporal order of occurrence of confounder and exposure, and their interdependence in time. When estimating the parameters of Model (2.10),  $A_{-1}$  and  $L_{-1}$  for  $k = 0$  are set to zero without loss of generality.

If there are more than two time intervals, the researcher should seek expert opinion on data-generating process to choose how many lags of past treatment and confounder history to include in the models for exposure and censoring. For the sake of generality  $\bar{A}_{k-1}$  and  $\bar{L}_{k-1}$  are used, which represent the whole treatment and confounder history.

For subject  $i$  the weights are computed as follows:

$$\widehat{SW}_i = \prod_{k=1}^2 \frac{[(\hat{p}_{i,k}^*)^{(A_{i,k})}] \times [(1 - \hat{p}_{i,k}^*)^{(1-A_{i,k})}]}{[(\hat{p}_{i,k})^{(A_{i,k})}] \times [(1 - \hat{p}_{i,k})^{(1-A_{i,k})}]}, \quad (2.11)$$

where  $p_k^* = P(A_k = 1 \mid A_{k-1})$ .

MSMs can also be used to estimate joint causal effects of two or more treatments on an outcome. This means that the exposure is multivariate. Let us consider an extension of Example 2.1, where chemotherapy cycles can both be delayed and can have the dose reduced. We denote the two components of the exposure in cycle  $k$  with  $A_k^1$  and  $A_k^2$ , respectively. The weights to be estimated are

$$SW_i = \prod_{k=0}^1 \frac{f[A_{i,k}^1, A_{i,k}^2 \mid \bar{A}_{i,k-1}^1, \bar{A}_{i,k-1}^2]}{f[A_{i,k}^1, A_{i,k}^2 \mid \bar{A}_{i,k-1}^1, \bar{A}_{i,k-1}^2, \bar{L}_{i,k}]}. \quad (2.12)$$

The numerator and denominator of Equation (2.12) could be estimated by using a bivariate logistic odds-ratio model (available in the R package `VGAM` [11]) if  $A_k^1$  and  $A_k^2$  are both binary variables. Alternatively, one could apply the definition of conditional probability and estimate  $SW_i$  as follows:

$$SW_i = \left[ \prod_{k=0}^1 \frac{f[A_{i,k}^1 \mid \bar{A}_{i,k-1}^1, \bar{A}_{i,k-1}^2]}{f[A_{i,k}^1 \mid \bar{A}_{i,k-1}^1, \bar{A}_{i,k-1}^2, \bar{L}_{i,k}]} \right] \left[ \prod_{k=0}^1 \frac{f[A_{i,k}^2 \mid \bar{A}_{i,k-1}^1, \bar{A}_{i,k-1}^2]}{f[A_{i,k}^2 \mid \bar{A}_{i,k-1}^1, \bar{A}_{i,k-1}^2, \bar{L}_{i,k}]} \right]. \quad (2.13)$$

When deciding which phenomenon will form the second component of the exposure, one has to take into account the temporal order of occurrence of the exposures. In the context of the example above, first we observe whether a cycle was delayed, and then whether its dose was reduced. That is why we can condition the administration of dose reduction in cycle  $k$  on delay of cycle  $k$ . On the contrary, if we condition cycle delays on dose reduction, it will be wrong, since we will be conditioning on the future.

In Section 12.6 [12], Hernán and Robins describe how to use the concepts of censoring and IPTW to recover data from records with missing end-of-study outcome. The

missing outcome can be seen as a censored observation. This procedure can be used in Example 2.1. Patients with missing HRe are treated as censored observations. IPCWs are estimated using all observations, i.e. including the records with unknown HRe. The main analysis is performed on the subset with observed outcome while weighting these observations with the estimated in the previous step subject-specific IPCWs. In this way possible selection bias due to sub-group analysis is overcome.

Finally, one might think that another way to correct for time-dependent confounders is to include them as covariates in Models like (2.7) and (2.9). Such models will fail to provide unbiased estimates of the causal effect of confounded time-dependent treatment since the treatment itself can affect subsequent covariate value and, therefore, should not be corrected for [4, 2]. Rosenbaum [13] shows that if treatment assignment process can be ignored given past treatment and confounder history, a bias is introduced when the time-dependent confounder is included as a covariate in the model. He decomposes the bias into two components. The first part is due to estimation of conditional population mean of the outcome given the exposure and the confounders in the sample. The second part is due to the inappropriate averaging of a conditional expectation to obtain a marginal expectation. The latter is due to the feedback between exposure and confounder. This issue in case of a binary outcome is further elaborated on in [14].

### 2.2.2 Assumptions

The validity of MSMs depends on a number of assumptions. The first one, *consistency*, was stated at the beginning of Section 2.2, and is recalled here for completeness of the exposition.

Consistency, in the context of MSMs, addresses the correspondence between factual and counter-factual outcomes.

**Assumption 2.1. Consistency** Let  $Y$  be the outcome of a given patient allowing treatment trajectory  $\bar{A}$ . Consistency assumes  $Y = Y_{\bar{a}}$  if  $\bar{A} = \bar{a}$ , where  $Y$  is the random variable corresponding to the factual outcome,  $Y_{\bar{a}}$  denotes the counter-factual outcome of treatment history  $\bar{a}$ .

MSMs rely on the well-known, untestable assumption of *no unmeasured confounding*. This assumption can be formulated as follows. If all common causes of the counter-factual outcomes and exposure at time  $k$  are recorded, there is no unmeasured confounding.

**Assumption 2.2. No unmeasured confounding** If for  $\forall k Y_{\bar{a}} \perp\!\!\!\perp A_k \mid \bar{A}_{k-1}, \bar{L}_k$ , then there are no unmeasured confounders given the data on measured confounders.

MSMs require the *exchangeability* assumption to hold, which implies lack of unmeasured confounding [5]. Exchangeability implies that counter-factual outcomes do not depend on observed but on potential exposure.

**Assumption 2.3. Exchangeability** Potential outcomes  $Y_{\bar{a}}$  are independent of observed exposure if  $f(A_k \mid \bar{A}_{k-1}, \bar{L}_k, Y_{\bar{a}}) = f(A_k \mid \bar{A}_{k-1}, \bar{L}_k)$ .

Exchangeability is an untestable assumption but one can study the sensitivity of the results under violations of this assumption [5]. Details about the sensitivity analysis procedure are outlined in [15].

Further, exposure and covariate history up to time  $k - 1$  should not restrict the domain of the exposure at time  $k$ . That is, *positivity* assumption should be met.

**Assumption 2.4. Positivity**  $P(A_k = a_k \mid \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k = \bar{l}_k) > 0$ , for all  $a_k, \bar{a}_{k-1}$ , and  $\bar{l}_k$ .

If  $L_k$  is a vector of time-dependent confounders, *random zeros* will likely occur in the data, i.e. we will not observe every value of  $A_k$  within every combination of  $\bar{L}_k$  and  $\bar{A}_{k-1}$ . This is not a violation of the positivity assumption but could still introduce bias and increase the variance of our estimates. As one wants to correct for all possible confounders, there is a trade-off between bias due to confounding and non-positivity [5]. Cole and Hernán [5] show that to use many covariates introduces bias due to non-positivity and at the same time increases the variance of the estimated effects. The same authors suggest that it is possible to truncate the weights to achieve a mean weight of 1 (see assumption of correct model specification below) and increase the precision of the estimates. However, caution is necessarily since truncation introduces bias.

Before estimating the causal effect of a time-dependent exposure in the presence of time-dependent confounding on an outcome through a MSM, one has to develop a model for allocation of the exposure and a censoring model. The validity of the inference is conditioned on *correct specification* of all models involved.

**Assumption 2.5. Correct model specification** 1) the structural model; 2) the exposure model; 3) the censoring model are correctly specified.

A necessary condition for correct model specification is that the stabilised weights,  $SW$  and  $SW^\dagger$ , have mean equal to one [4, 5]. This assumption is the cornerstone of MSMs. It is widely recognised that construction of causal diagrams through DAGs together with field experts can guide model building.

## 2.3 Applications of MSM

This section is a review of published articles which show the use of MSMs in practice. The first two examples cover binary outcomes although most applications of MSMs use survival data and construct models for time-to-event outcomes.

Bodnar *et al.* [16] built a MSM to estimate the causal effect of iron supplementation during pregnancy on odds of anaemia at delivery, where iron supplementation is treated as a time-varying exposure. The authors address the bias of crude analysis and traditional adjustment for confounding in the context of time-dependent confounding. Initial iron treatment was randomly allocated within strata defined by baseline serum ferritin and haemoglobin level, and could be modified once, at 24-29 weeks of gestational age, based on current values of the same blood markers of iron status and treatment side effects.

Lancia *et al.* [6] developed a MSM to estimate the joint causal effect of dose reductions of methotrexate and chemotherapy-cycle delays on HRe, a dichotomous outcome. They use the preoperative data from a randomised clinical trial in osteosarcoma. The time-dependent confounding lies in the negative feedback between therapy modifications, in the form of dose reductions and cycle delays, and the toxicities due to therapy. Further, the authors describe the highly complex structure of longitudinal chemotherapy data and how this aspect is taken into account in the MSM. Practical issues concerning the estimation of IPTWs and R packages are discussed.

Hernán *et al.* [17] used a MSM to estimate the causal effect of Zidovudine therapy, a dynamic treatment regimen [18], on mean CD4 count among HIV-infected men in the Multicenter AIDS Cohort Study. The novelty of their study consists of application of MSMs on observational repeated measurements data. They used up to 16 visits per patient with waiting period of 6 months. In the same article the authors simulated sequentially randomised treatment based on time-dependent covariates through the use of IPTWs within the framework of Generalised Estimating Equations. The authors also discuss differences of estimation methods in case treatment is randomised at baseline or non-randomised treatment assignment as in an observational study.

Using the same dataset, Hernán *et al.* [7] estimated the causal effect of Zidovudine on survival of HIV-positive men through MSCPHM. CD4 count is a risk factor for mortality and a predictor of subsequent initiation of zidovudine therapy, i.e. it is a time-dependent confounder of the effect of Zidovudine on survival. The effect of time-dependent covariates, i.e. Zidovudine therapy and CD4 count, on the outcome is estimated by employing a Cox regression model. For every patient a weight is estimated at every visit, which incorporates the treatment and confounder history. The weights, used to construct a pseudo-population where treatment is no longer confounded, are also time-dependent. Time-specific weights option is not yet available in statistical packages that support Cox models. Instead, the authors used pooled logistic regression where every patient-visit is treated as an observation. The effect of time is incorporated through a time-specific model intercept. The equivalence of Cox proportional hazards model and pooled logistic regression is discussed in [8, 9].

In 2001 Hernán, Brumback, and Robins [2] extended the MSMs to estimation of joint causal effects. For the analysis data from a multicenter observational study of homosexual men are used. The joint causal effect of Zidovudine therapy and prophylaxis therapy for *Pneumocystis carinii* pneumonia on survival is investigated through a MSCPHM. Both treatment effects are confounded by CD4 count, a time-dependent covariate. They illustrate the joint causal effects methodology in the context of binary and time-to-event outcome, and describe the process of estimation of IPTW.

Recently, MSMs have been used in a number of studies to estimate joint causal effects on time-to-event outcomes [19, 20]. Both articles devote special attention to the construction of the models for the weights and their estimation.

As shown by Hernán *et al.* [18], MSMs can be used to compare treatment regimens. The methodology is as follows: two regimens of interest must be defined; artificially censor individuals when they stop following one of the regimens of interest; estimate

inverse probability weights to adjust for potential selection bias introduced by censoring in the previous step; compare the outcome of interest of the uncensored individuals under each regimen of interest by fitting a weighted regression model with dichotomous regimen indicator and baseline confounders as covariates.

## 2.4 Alternative methods for causal inference

This section presents a short review of two alternative methods for causal inference when time-dependent confounding is present: the mini-trials approach and Structural Nested Model (SNM). At the end of the section we motivate our choice of MSMs over the others methods.

A surrogate for MSMs is the mini-trials approach in survival analysis [21, 22]. Patients enter the trial at different time. These times, together with the times when treatment is modified, divide patients' follow-up time in fixed-exposure intervals. At the beginning of each of these intervals a new fictitious two-arm mini-trial starts. A patient takes part in as many mini-trials as the number of intervals he is followed up. Patients are followed until the occurrence of event of interest, end of follow-up, or exposure modification, where time takes discrete values measured in intervals. Exposure modification forces artificial censoring. If a patient switches to the other regimen of interest he is "randomised" to the other arm for the next mini-trial, which starts from the next time interval. End of follow-up introduces non-informative censoring also known as administrative censoring, but exposure modification is a selective form of artificial censoring, which introduces bias. This bias is corrected for by using inverse probability-of-censoring weighted estimators.

SNM and Structural Nested Failure Time Model were introduced prior to MSMs to model the degree to which the effect of current treatment is modified by past treatment and past time-dependent confounder history [23, 24, 25, 26]. These methods are a form of additive models. They model the effect of the time-dependent exposure on the end-of-study counter-factual outcomes separately for each time period. The focus lays on estimating the mean change in the outcome due to exposure in each time period. This is possible due to the additive decomposition property of conditional outcome mean. The name of the methods stresses the nested nature of the time-specific sub-models. The parameters of these models are estimated through the use of a technique known as  $g$ -estimation. Most of the assumptions of MSMs also apply for SNMs.

We have decided to use MSMs over the mini-trials approach because the latter is limited to applications with very simple, one-dimensional exposures. MSMs closely resemble standard regression models, and SNMs cannot be used to estimate the effect of a treatment on dichotomous outcomes unless the outcome is rare. This is because logistic SNMs cannot be fit by  $g$ -estimation. In 1999 Robins [3] showed that MSMs, in contrast to SNMs, cannot estimate the effect of time-dependent exposure on an outcome when there are values of the time-dependent confounder for which the exposure is predetermined. This is a special case of violation of the positivity assumption. The primary goal of SNMs is to explore the ability of the time-dependent confounder to modify the effect

of the exposure on the outcome. MSMs can only assess interactions between baseline confounders and the time-dependent exposure [3]. Further, SNMs can be consistently estimated in presence of unmeasured confounding if data on an instrumental variable can be obtained [2, 3]. Finally, in presence of strong covariate-treatment associations, theoretical arguments imply that it should be possible to construct confidence intervals based on  $g$ -estimation of SNMs that are both narrower and have better coverage properties than those based on IPTW estimation of MSMs [7].

## Chapter 3

# Data description

The clinical background needed to build the model to address the research question and the corresponding dataset are described in detail in this chapter. In Section 3.1 a general introduction to the data and the data gathering process is discussed. It is followed by a descriptive statistics section. Section 3.3 is devoted to the exposures of interest and describes their distribution. In Section 3.4 the interplay between exposure and confounders is described. The chapter ends with a short discussion of missing values.

### 3.1 Data gathering process

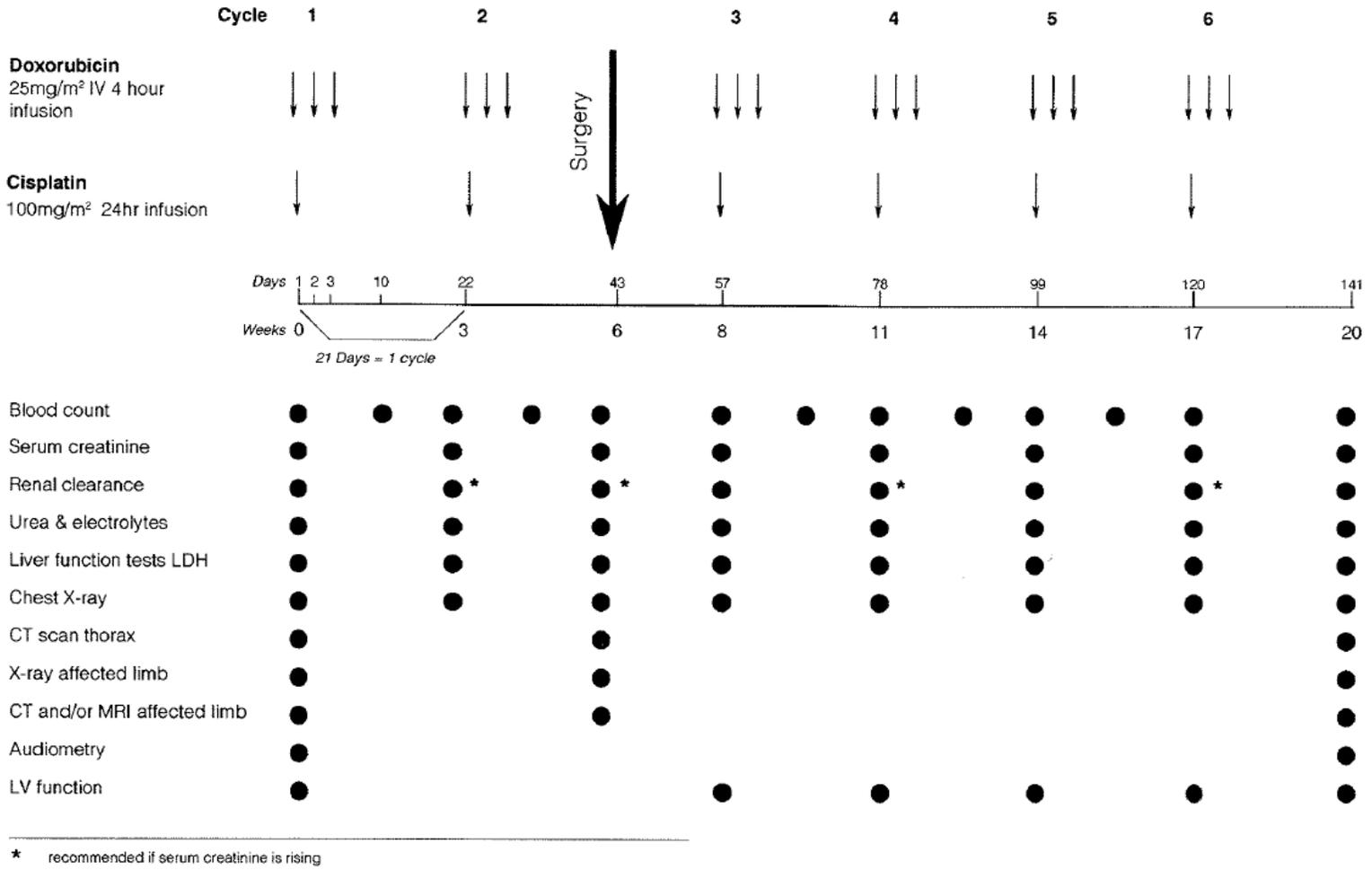
In August 1993 the European Osteosarcoma Intergroup (EOI) initiated a phase III randomised open-label multi-center clinical trial of chemotherapy in osteosarcoma (study number EORTC 80931/MRC BO06 [27]). The aim of the trial was to evaluate potential benefit from increasing the planned dose intensity of cytotoxic drugs on patients' survival. Although higher dose intensity did not prolong survival, a significant association between Good (histological) Response (GR) and high received dose intensity was found [28]. Previous EOI trials that used two-drug chemotherapy regimen of CISPLATIN (CDDP) and DOXORUBICIN (DOX) as the conventional arm showed no evidence of benefit when the number of agents or the length of treatment were increased [29]. On the contrary, it has recently been shown that an addition of a third drug to a two-drug regimen significantly improves survival of osteosarcoma patients but the same does not apply to a fourth drug [30].

The aforementioned randomised study addressed the hypothesis that survival could be improved by increasing the planned dose intensity of CDDP and DOX. This was done by contrasting the two-drug regimen of six chemotherapy *cycles* each with a nominal length of three weeks to a two-week cycles regimen. The contraction of chemotherapy courses was made possible by additional infusion of Granulocyte Colony-Stimulating Factor (G-CSF). Patients aged 40 years or less with operable osteosarcoma in an extremity long bone were recruited for this trail from August 1993 until October 2002. Patients' diagnosis was confirmed with biopsy, an analysis of a sample of the modified tissue. Up to four weeks since biopsy, the patients who met the eligibility criteria were randomised

EORTC 80931

REGIMEN 1 (without G-CSF)

MRC BO06



32

Figure 3.1: Control arm treatment and monitoring schedule

to follow one of the two treatment regimens of interest.

The goal of the analyses in this thesis is to assess the causal impact of therapy modifications on postoperative and survival outcomes. The problem is addressed using only the records from the conventional arm with standard cycle duration. Since the main component of therapy modifications rests upon cycle durations, the two arms cannot be analysed simultaneously.

The treatment in the arm of interest consists of two preoperative chemotherapy cycles, surgery to resect the tumour with limb salvage wherever possible, and four postoperative chemotherapy cycles. The treatment schedule is presented in Figure 3.1. Treatment timing plays a central role for the successfulness of the intervention, and therefore requires close co-operation between oncologists and surgeons.

Each chemotherapy cycle starts with a 24-hour infusion of  $100 \text{ mg/m}^2$  CDDP, and four-hour infusion of  $25 \text{ mg/m}^2$  DOX during the first three days. Then a recovery period of 18 days follows. The two drugs target different stages of the cell reproduction process, and damage their division process in the attempt of triggering apoptosis (programmed cell death). As such chemotherapy is most effective on cells that are actively reproducing (e.g. bone marrow and tumour). In addition to damaging tumour cells, chemotherapy harms normal body cells in different tissues and their processes. The severity of clinical side effects is measured by CTCAE toxicity grades [31]. Possible grades range between 0 and 4 corresponding to increasing severity. Since the negative effect of chemotherapy on different organs could disable their function and cause an adverse event, body response to chemotherapy is monitored throughout the cycles.

The tracking of physiological functions in a human body is performed in the following way. Before the start of each chemotherapy cycle ten laboratory values are measured, namely White blood cell (WBC) count, neutrophils, platelets, renal clearance, alkaline phosphatase, lactate dehydrogenase, calcium, magnesium, liver function tests, and electrolytes. Their rate should exceed certain thresholds in order to allow a cycle to begin. Three of the aforementioned substances (WBC, neutrophils and platelets) are also recorded on day 10 and 22 of each cycle. Treatment side effects, namely nausea, oral toxicity, infection, neurological and cardiac toxicity, and ototoxicity are given a CTCAE grade throughout each course. All this information is recorded cycle-wise using chemotherapy form displayed in Figure 3.2. In addition, cardiac function is monitored through echo-cardiogram prior to treatment, before cycles 3, 4, 5, and 6, and after 3 and 12 months since completion of chemotherapy.

**EOI TRIAL OF CHEMOTHERAPY +/- G-CSF IN OPERABLE OSTEOSARCOMA  
CHEMOTHERAPY FORM**

Please complete this form at the end of each cycle of chemotherapy and send top copy to your data coordinating centre.

Patient's name/ID code: .....       Date of birth (D,M,Y)  
Institution: .....     EORTC/MRC patient number

1     Cycle number  
2       m<sup>2</sup> Surface area  
3     cm Height  
4     Kg Weight

5       Date of start of cycle (D,M,Y)  
6     mg CDDP Dose this cycle  
8     mg DOX Total dose this cycle

7  Major reason for delay/reduction  
1 = No delay/reduction  
2 = Administrative  
3 = Haematological toxicity  
4 = Renal toxicity  
5 = Ototoxicity  
6 = Neurotoxicity  
7 = Infection  
8 = Other, specify .....

9  Major reason for delay/reduction  
1 = No delay/reduction  
2 = Administrative  
3 = Haematological toxicity  
4 = Renal toxicity  
5 = Ototoxicity  
6 = Neurotoxicity  
7 = Infection  
8 = Other, specify .....

**FOR G-CSF REGIMEN ONLY**  
10       µg G-CSF Total dose this cycle  
11  Was any toxicity attributed to G-CSF  
1 = No  
2 = Yes, specify .....

Laboratory values	Before start of cycle	Day 8 (G-CSF) or Day 10	Day 15 (G-CSF) or Day 22
Date of test (D,M,Y)	12 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	23 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	27 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
WBC x 10 <sup>9</sup> /L	13 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	24 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	28 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Neutrophils x 10 <sup>9</sup> /L	14 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	25 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	29 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Platelets x 10 <sup>9</sup> /L	15 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	26 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	30 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>
Renal clearance ml/min/1.73m <sup>2</sup>	16 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
Alkaline phosphatase IU/L	17 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
LDH IU/L	18 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
Calcium mmol/l	19 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
Magnesium mmol/l	20 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
LFT	21 <input type="text"/>		
Electrolytes	22 <input type="text"/>		

**Side effects of treatment (Please use toxicity coding on back of book and specify grade)**

31  Nausea/Vomiting  
32  Oral  
33  Infection  
34  Other, specify .....

35  Neurological  
36  Ototoxicity  
37  Cardiac

Signed ..... Date .....

Figure 3.2: Chemotherapy form  
34

## 3.2 Descriptive statistics

The description in the previous section is meant to introduce the reader to the clinical background and the multidimensionality of the records. The present section visualises and describes the available records.

Among patients who were randomised to the control arm, 218 patients in total meet the trial eligibility criteria. Figure 3.3 presents the sample distribution of the two key demographic patient characteristics: age and gender. Both are distributed as described in the clinical literature of osteosarcoma. The studied subjects were between 4 and 38 years old with peak during adolescence. Males develop the disease slightly more often than females (58.6% and 41.4%, respectively).

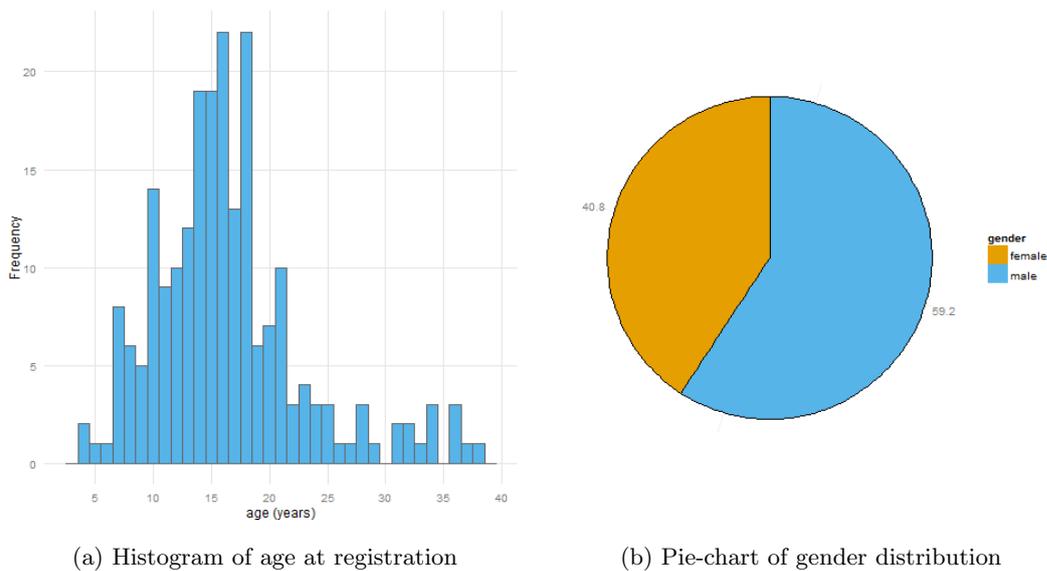


Figure 3.3: Plots of key demographic variables

Although the protocol states that six chemotherapy cycles should be given only 80.7% of the patients complete all six cycles. Among the remaining 19.3%, about 40% abandoned the study after two cycles of chemotherapy only. Figure 3.4 shows the distribution of drop-outs per cycle and the associated reasons to terminate treatment. Treatment discontinuation at the end of the second cycle is due to tumour progression or surgery refusal (which is planned after cycle 2). These are shown in the orange and green sections in Figure 3.4b. The largest fraction of treatment refusals is observed among those who complete only two cycles. Excessive toxicity is the major reason to terminate treatment in later cycles. Throughout cycles 1 to 5, there are patients who discontinue treatment due to reason other than the listed options (category *other*).

Surgery is a key component of the therapy. As stated earlier, treatment timing is very important, since delays allow damaged cells to recover and tumour to regrow. Figure 3.5 presents the two aspects of surgery timing - after which cycle the operation is performed

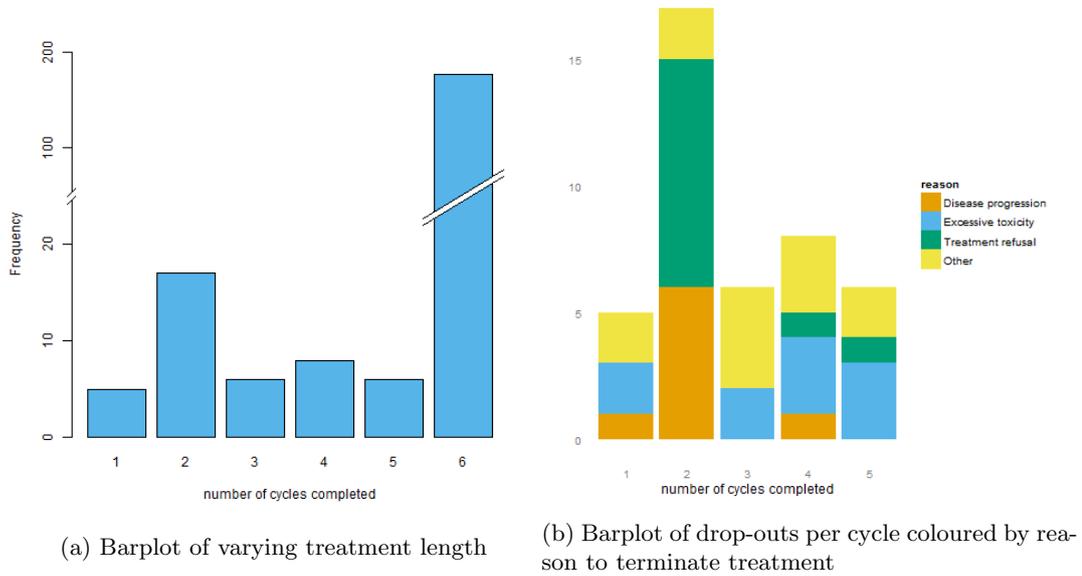


Figure 3.4: Plots of protocol treatment completion

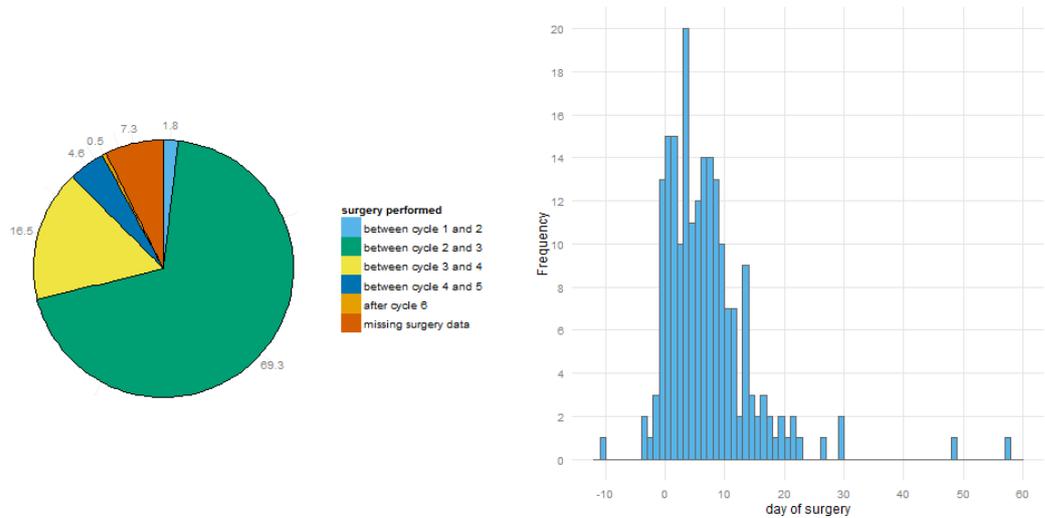
(Figure 3.5a), and how long it took for surgery to take place after completion of the last preoperative chemotherapy cycle (Figure 3.5b). In the interest of presentation, Figure 3.5b omits one record with surgery on day 130 since completion of last preoperative cycle. Only 69.3% of the patients were operated between cycle 2 and three, according to the protocol. A small fraction were resected prematurely, most likely because of disease progression, while 7.3% of the patients lack surgery information. Most patients were operated within 10 days after completion of the last preoperative cycle. Surgery form of the trial states that surgery is performed *on time* if a patient undergoes surgery within  $\pm 2$  days of the end of the second preoperative cycle, i.e. between day 41 and 45 of the therapy.

The outcomes of interest are defined as follows. Histological response (HRe) is a dichotomous variable indicating whether the percentage of viable tumour after resection exceeds 10%. Less or equal to 10% viable tumour is declared a *good histological response* (GR), while more than 10% is considered a *poor histological response* (PR). The distribution of this variable is visualised on Figure 3.6a. Only 29% of the patients achieve GR, while 17% lack record of HRe. Some of the patients with missing information about HRe did not receive any surgery (see Figure 3.5a). The remaining patients were recorded a PR.

**Definition 3.1. Histological response** *Good histological response* corresponds to  $\leq 10\%$  viable cells in the resected specimen.

**Definition 3.2. Overall Survival:** the event of interest is death due to any cause.

**Definition 3.3. Event-Free-Survival:** the event of interest is disease progression,



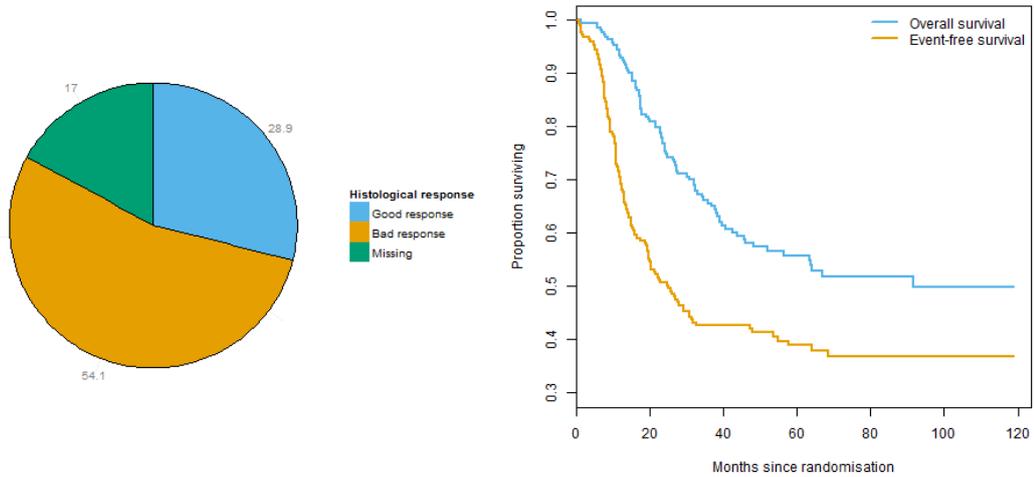
(a) Pie-chart of surgery time in cycles (b) Histogram of day of surgery since completion of last preoperative cycle

Figure 3.5: Plots of surgery timing in cycles and in days since completion of the last preoperative cycle

local recurrence, distant metastases, or death due to any cause, whichever comes first.

Some patients in the trial were followed for almost 10 years since randomisation. Figure 3.6b presents Kaplan-Meier estimates of Overall Survival (OS) and Event-Free-Survival (EFS) functions in months since randomisation. A lot of patients experience any form of disease progression or recurrence since EFS curve is below the OS curve. There are 5 patients who experience disease progression during therapy. About 25% of the patients do not complete the whole six cycles chemotherapy, and nominal study time (20 weeks) is too short compared to follow-up time, the Kaplan-Meier estimates of survival curves start to drop from the very beginning. According to the protocol, therapy should end within 20 weeks. By that time one patient has died, and 8 have experienced an adverse event (disease progression, recurrence or metastases). The death rate seems constant from month 5 to 40. Mean OS and EFS are estimated to be 74.74 (3.48) and 55.00 (3.60) months, respectively, where standard errors are given in brackets.

The treatment protocol suggests that haematological, oral, renal, and cardiac toxicities, ototoxicity and infection are the most important treatment side effects and therefore they should be carefully monitored. Renal toxicity is measured only once at the start of each cycle, when all laboratory measures should allow chemotherapy to be given. It was not recorded for more than 30% of the patient-cycles. This implies that we cannot address the impact of renal toxicity on the outcomes of interest. Figure 3.7 lists CTCAE grades corresponding to other toxicities, divided per cycle. The plots in Figure 3.7 show the study drop-out rate by visualising the number of patients that complete each subsequent cycle. In addition, they show the variation in the sample distribution

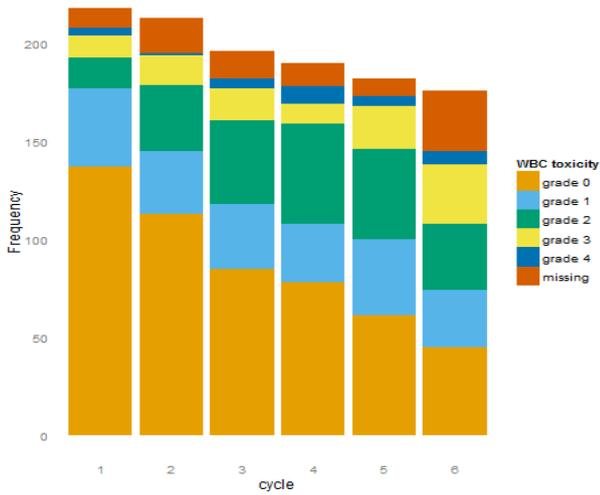


(a) Pie-chart of histological response

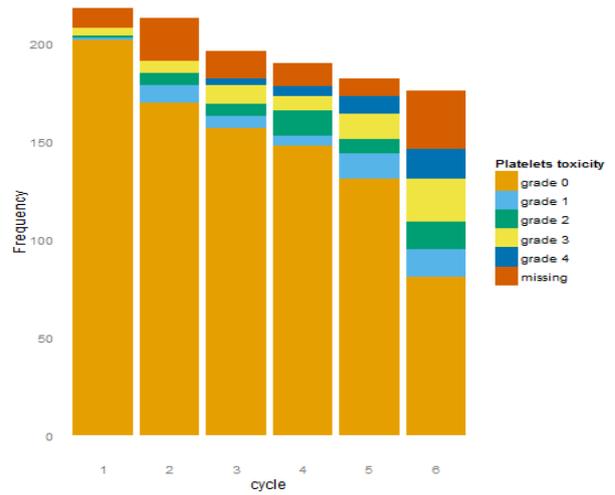
(b) Kaplan-Meier estimates of overall and event-free survival functions

Figure 3.6: Plots of histological response, overall and event-free survival

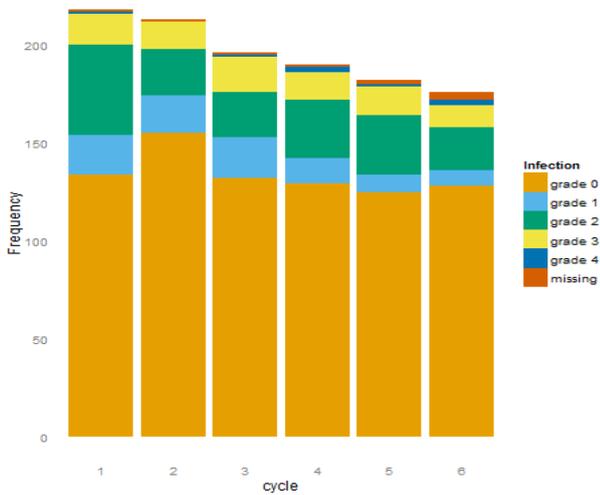
of side effects measured with CTCAE grades. Plots (a) and (b) in Figure 3.7 refer to haematological toxicity, also referred to as *myelosuppression* (WBC and platelets are both types of blood cells). According to the protocol, their count is recorded on day 22 of each cycle, i.e. at the nominal end of each cycle. After that the record is translated into a CTCAE grade. These values are missing in 8% of the patient-cycles. WBCs exhibit largest variability among all side effects. Infection and the other organ toxicities were assigned a CTCAE grade by the oncologist based on patient's condition throughout the cycle. Relatively constant number of patients, about 135, had no infection during at least one of the cycles. Grade 4 is hardly ever assigned and there are almost no missing values. The distribution of oral toxicity grades from cycle to cycle is relatively constant. This is unexpected because clinicians believe that higher grade toxicities appear more frequent in late cycles. Cardiac and ototoxicity, both late toxicities, are highly invariant and have a constant fraction of missings.



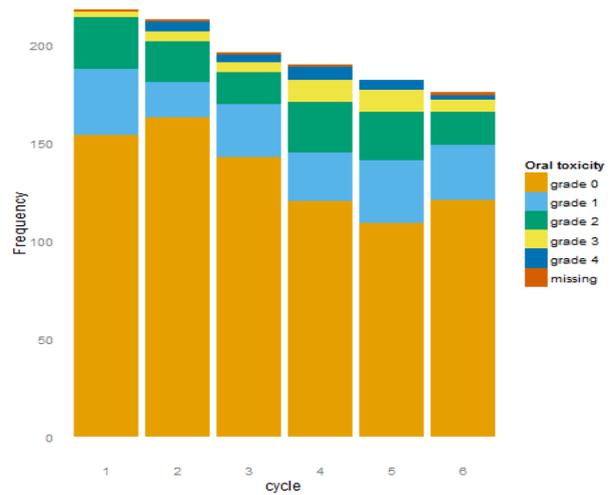
(a) CTCAE grades of WBC per cycle



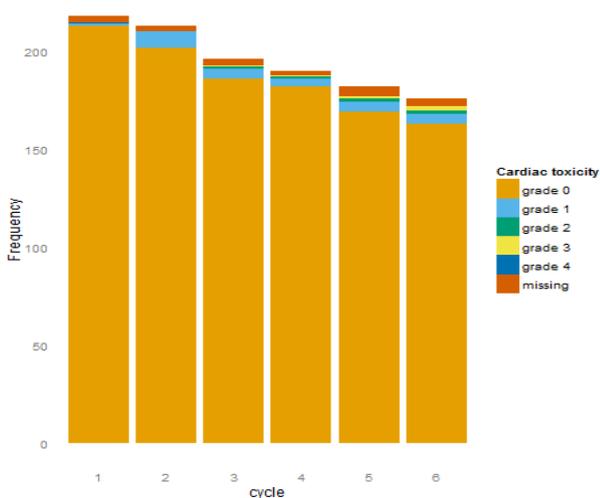
(b) CTCAE grades of Platelets per cycle



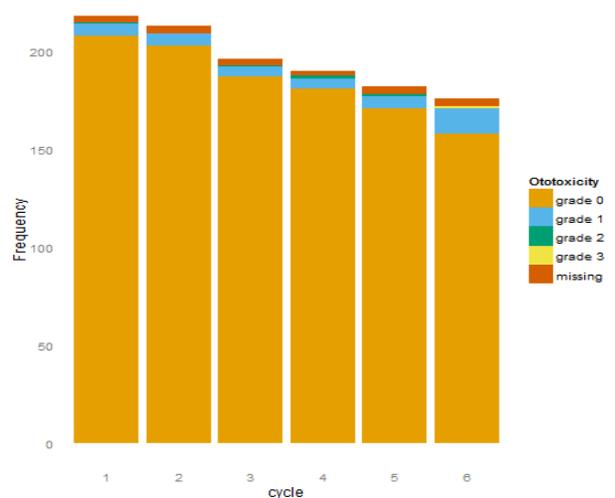
(c) CTCAE grades of Infection per cycle



(d) CTCAE grades of Oral toxicity per cycle



(e) CTCAE grades of Cardiac toxicity per cycle



(f) CTCAE grades of Ototoxicity per cycle

Figure 3.7: Barplots of CTCAE grades per toxicity and cycle

### 3.3 Exposures of interest

Besides nominal doses and timing, the study protocol prescribes therapy modifications based on treatment side effects and the general state of a patient. Therapy modifications consist of dose reductions/discontinuations and treatment delays. The analyses performed in this thesis are meant to estimate causal effects of therapy modifications on HRe and patients' survival. That is why the various therapy modifications comprise the exposures of interest in the analyses that follow.

In terms of chemotherapy-cycle delays, the study protocol prescribes:

- one week delay (or less in case of full recovery) when WBC or platelets count correspond to CTCAE grade 2 or larger,
- delay until full recovery if oral toxicity occurs.

Study protocol prescribes either 20% dose reduction for all subsequent courses in case of:

- joint occurrence of myelosuppression and infection,
- oral toxicity grade 3 or 4,

or discontinuations of:

- CDDP in case of ototoxicity,
- DOX in case of cardiac toxicity.

Based on the guidelines in the protocol and published analyses on this dataset [28], exposures of interest are defined in the following way:

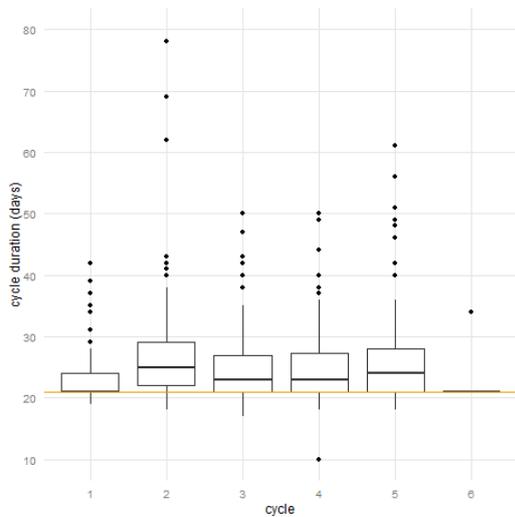
**Definition 3.4. Cycle delay** A cycle is delayed if it starts later than day 23 of the preceding cycle, i.e. in case of delay  $\geq 3$  days.

**Definition 3.5. Dose reduction** Dose reduction occurs if received dose of at least one of the two drugs is  $\leq 86\%$  of the nominal dose.

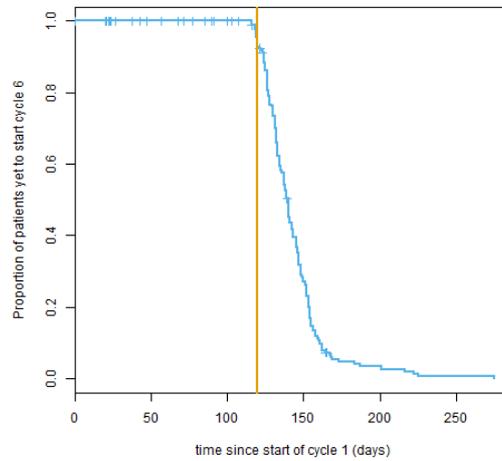
In order to identify delays, we first need to calculate *cycle duration*. Cycle  $k$  duration is equal to:

- starting date of cycle  $k + 1$  minus starting date of cycle  $k$ ;
- date of surgery minus date of start of cycle  $k$  if cycle  $k$  is last preoperative cycle;
- 21 days, if cycle  $k$  is patient's last chemotherapy cycle.

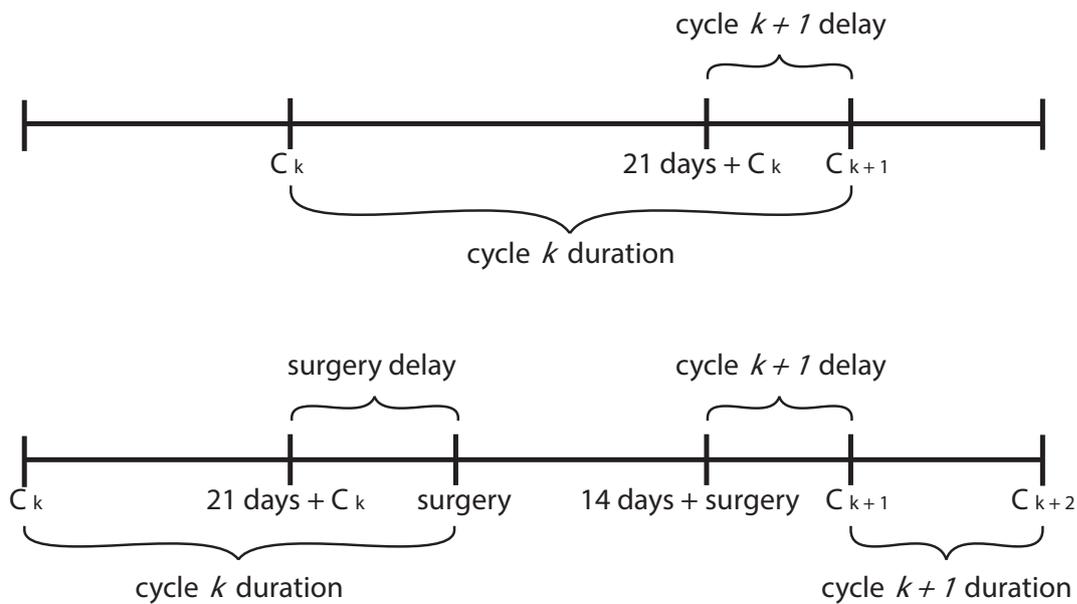
As discussed above, in order to determine if a cycle was delayed we need to know the date of surgery. This information is not available for 7.3% of the patients. Those patients are therefore excluded from the analyses. The graphs that follow, as well as



(a) Boxplots of cycle duration



(b) Kaplan-Meier estimate of time to begin cycle 6



(c) Delay calculation scheme

Figure 3.8: Plots of different aspects of treatment duration

plot (b) in Figure 3.5, are produced using the records of 202 patients with known date of surgery.

Figure 3.8a visualises the distribution of cycle durations per cycle number. Throughout all chemotherapy courses, there are patients whose cycle lasted more than the nom-

inal length of 21 days (see orange horizontal line). The longest durations are attributed to the second cycle since most of the surgeries were performed between cycle 2 and 3. As shown on Figure 3.8c, the duration of the last preoperative cycle incorporates surgery delay. Cycle 6 cannot last longer than 21 days since it is the last cycle for each patient. The only exception is the patient with surgery performed after cycle 6. There are also patients with shorter cycles. These are most probably preoperative cycles. There is one outlier whose fourth cycle was only 10 days long. It would be that this date of surgery was incorrectly recorded.

In order to show how long it takes to complete therapy from another viewpoint, Figure 3.8b presents a Kaplan-Meier estimate of time to begin cycle 6. It was computed on all patients, i.e. 218, and those who discontinue treatment before starting cycle 6 are censored. The orange vertical line indicates the time to start cycle 6 according to the protocol, i.e. 120 days since start of cycle 1. Most of the patients with less cycles are censored before day 120. A limited number of patients begin their sixth course of chemotherapy prematurely. Mean time to start cycle 6 is 142.89 (1.63) days since start of cycle 1. This amounts to average chemotherapy course extension of 4.6 days (since cycle 6 cannot be extended, we divide the difference by five cycles). Almost all patients have started cycle 6 by day 175.

As stated above, we want to estimate the causal effect of cycle delays, a quantity that can be derived from cycle duration. In Figure 3.8c it is shown how to compute cycle delays. The horizontal lines represent the time axis. The top example displays the general case. The delay of cycle  $k + 1$  is the extension of the duration of cycle  $k$  beyond 21 days. However, this is not the case for the first postoperative cycle. Its reference is surgery recovery time. The protocol states that surgery should be performed on the 43rd day of the therapy. Thereafter a patient has two weeks to recover and the third cycle should start on day 57 of therapy. Leaving aside surgery timing in terms of cycles, first postoperative cycle is delayed if it starts later than day 14 after surgery. Note that cycle delays occur before a cycle has started. The surgery could be delayed if it does not start on day 22 of the last preoperative cycle. By definition, cycle 1 cannot be delayed. Therefore, there are at maximum five delayed cycles. As will be shown later, surgery delays occur mainly because of administrative reasons. Therefore, surgery delay will be regarded as an independent exposure.

Cycle  $k$  delay is equal to:

- in the general case, start of cycle  $k$  minus start of cycle  $k - 1$  minus 21 days, or
- if  $k$  is first postoperative cycle, start of cycle  $k$  minus date of surgery minus 14 days,

while surgery delay is equal to date of surgery minus start of last preoperative cycle minus 21 days.

Figure 3.9a shows the distribution of both types of delays (cycle and surgery). In the interest of presentation, three surgery delays longer than 40 days were omitted. Median surgery delay is 5 days, while cycles are delayed on average 2.8 days. Some cycle delays are negative. This occurs when the subsequent cycle starts in advance. According to

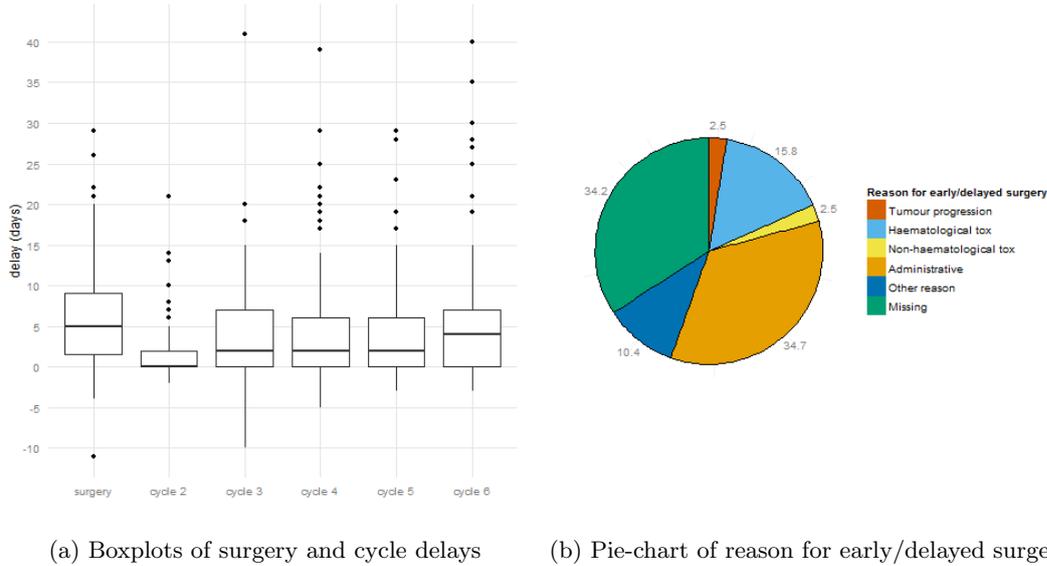


Figure 3.9: Plots of delays and associated reasons for delay

Definition 3.4, a cycle delay larger than 2 days is exposure of interest; this occurs in 35% of patient-cycles.

The study protocol leaves room for recording the reason for therapy modification. Since this field is common for cycle delays and dose reductions, it is hard to identify what the stated reason corresponds to. However, the reason for surgery delay was recorded separately. Its distribution is shown in Figure 3.9b. As stated earlier, the largest proportion of surgery delays is due to administrative reasons because performing a surgery on time requires close collaboration between surgeons and oncologists. The reason to delay surgery was not recorded for 34% of patients. This suggests that the surgery was performed on time since only six of these patients had surgery delay of a week or more.

**Definition 3.6. Surgery delay** A surgery is delayed if it starts later than day 27 of the last preoperative cycle, i.e. in case of surgery delay  $\geq 7$  days.

Next, we discuss the other exposure of interest – dose reduction. In order to understand the concept of dose reduction (Definition 3.5) we need to define *standardised dose*. The protocol prescribes doses in terms of milligrams ( $mg$ ) per  $m^2$  of body surface area. This measurement scale is motivated by body-surface-area's ability to indicate metabolic mass better than, for example, body weight. Metabolic mass expresses body's ability to process substances. Therefore, body surface area is used to determine suitable dosages. The nominal dose of DOX is  $75\text{ mg}/m^2$ , and that of CDDP is  $100\text{ mg}/m^2$ . Standardised dose is the ratio of administered dose to planned dose. If standardised dose is equal to 1, no reductions were applied. The protocol prescribes dose reductions of 20% but Figure 3.10 shows that this is not always the case in practice. Our definition of dose reduction

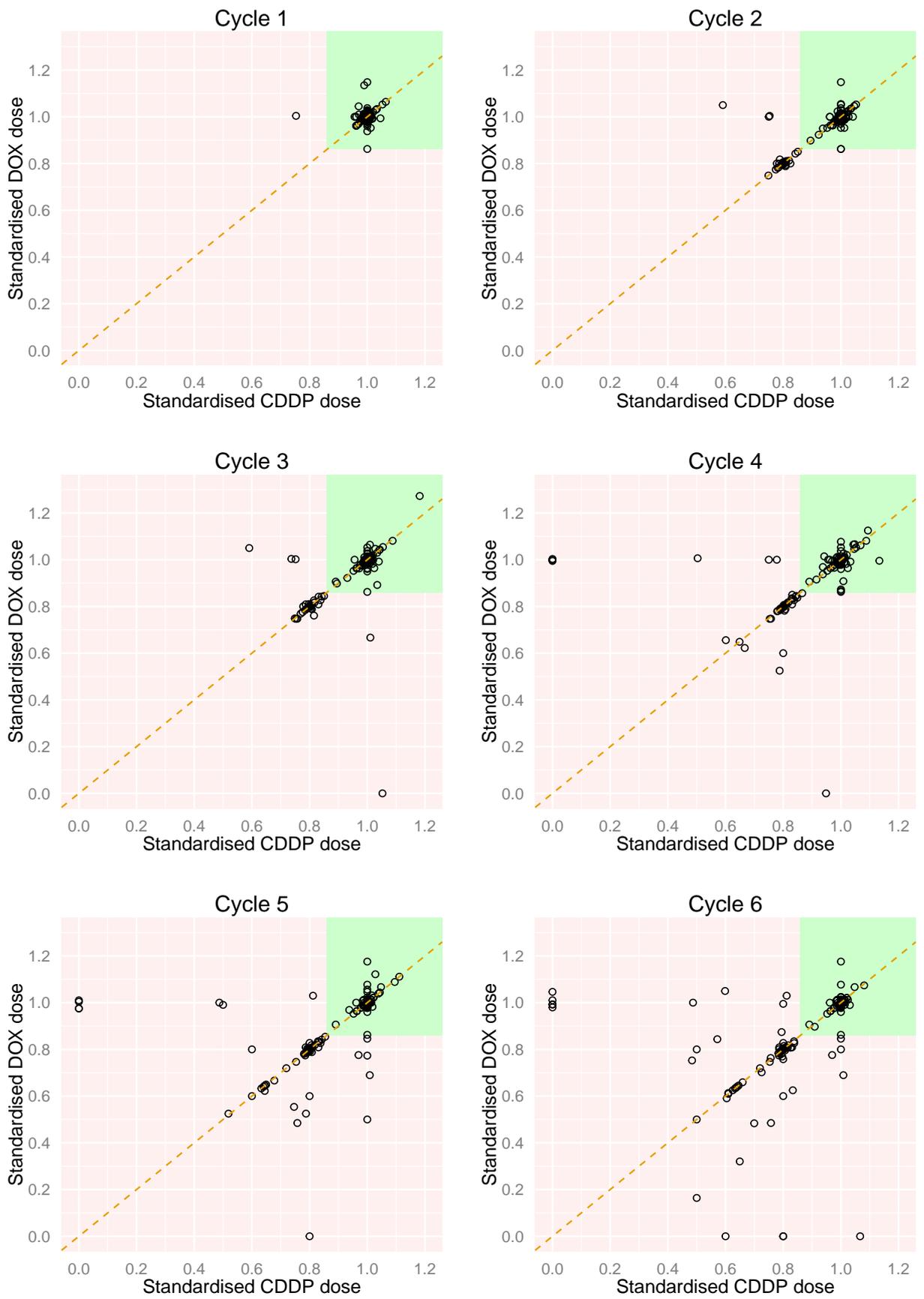


Figure 3.10: Scatterplots of standardised dose of CDDP vs. standardised dose of DOX per cycle

allows 7.5% tolerance around the expected reduction of 20% in order to capture all reductions, which vary around 80% due to rounding and other calculation errors. Mathematically standardised dose is defined as

$$\Delta = \text{standardised dose} = \frac{\text{dose administered mg/m}^2}{\text{dose planned mg/m}^2}. \quad (3.1)$$

Figure 3.10 presents standardised doses of the two drugs per cycle. The figure background is coloured in regions of *no dose reduction* (green area) and *dose reduction* (red area). The orange diagonal line marks reductions of the same proportion in both agents of the regimen. As one might expect, dose reductions as well as discontinuations more frequently occur at late cycles; the latter are indicated by circles that lie on one of the two axes. The first occurrence is discontinuation of DOX in cycle 3. In all plots on Figure 3.10 there is a clear separation between reduced and non-reduced doses. However, the reduced doses are not all equal to 0.8 or less. The clouds of points around the value of 0.8 on both axes motivate a tolerance of 7.5%. That is why the cut-off point for dose reduction is set to 0.86, i.e any standardised dose less than 0.86 is considered reduced. These variations in standardised dose are most probably driven by the lack of unified nomogram for calculation of body surface area, since a lot of nomograms can be found in the literature but there is none that accompanies the protocol.

### 3.4 Exposure-confounder feedback

The relationship between exposure, confounder and outcome is visualised in Figure 3.11. Chemotherapy is given to kill tumour cells. However, some normal body cells are also damaged. The extent to which this happens is recorded as toxicity grades of different tissues. Treatment side effects can cause an adverse event, and thus are associated with the outcomes. As stated earlier, study protocol lists guidelines for treatment modifications based on toxicities. Taking the last two facts together into account, it means that treatment side effects are confounders of the effect of exposure on the outcome.

The three key elements – exposure, confounder and outcome – are linked in the following way. Both exposure and confounder influence the outcome. However, they also interplay between each other. If toxicities occur, treatment is modified by applying dose reductions/discontinuations and/or cycle delays, i.e. presence of side effects triggers application of exposures of interest. These therapy modifications are meant to decrease toxicities and avoid potential adverse event. Hence, exposure decreases severity of the confounder. This effect circulation is referred to as *exposure-confounder feedback*.

Figures 3.12 and 3.13 try to visualise this phenomenon. Toxicities from cycle  $k$ , recorded throughout the cycle, modify exposure in cycle  $k + 1$ . Therefore, the exposures are presented on both axes per cycle starting from cycle 2, and coloured according to toxicities from the previous cycle, i.e. the colouring of cycle- $k$  sub-plot shows the confounders throughout cycle  $k - 1$  and the treatment-adjustment process.

As mentioned above, six types of treatment side effects are considered most important by clinicians. Figure 3.12 comprises of five plots. In each plot standardised dose of DOX+

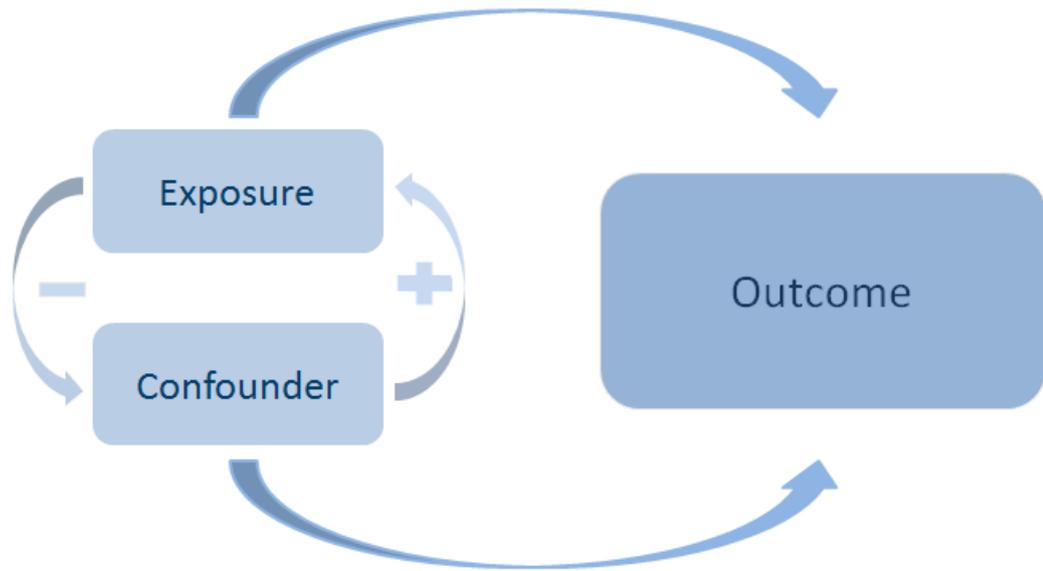


Figure 3.11: Relationship between exposure, confounder and outcome

CDDP, an average of standardised dose of each drug, is located on the horizontal axis. Cycle delays in days are expressed on the vertical axis. Every patient-cycle is a point in this two-dimensional space. The size of the marker displays the CTCAE grade of the severest toxicity experienced by the patient throughout the previous cycle. Larger circle corresponds to higher CTCAE grade. Colours are used to visualise the type of the most severe toxicity from the previous cycle.

In the plot for cycle 2, it is shown that platelets is often the severest toxicity. Nevertheless, their grade was not more than 2 in most cases. We see that the smallest circles are always red. If the severest toxicity has grade 0, this implies that all toxicities have grade 0, then colouring is not informative. In cycles three through six, the most severe toxicity is often WBC count.

If protocol had been carefully followed, one would expect to see larger circles anywhere around the plots except in the bottom right corner because this location corresponds to lack of delay *and* reduction. This would imply delays and/or dose reductions in the next cycle in case of severe toxicity. However, such a straightforward relationship is not evident from Figure 3.12. Nonetheless, this figure does not take into account the effect of multiple toxicities. That is why Figure 3.13 is also added.

Location of the points on the plots in Figure 3.13, as well as their sizes, are the same as in Figure 3.12. The only difference is that in Figure 3.13 the colouring is used to represent the number of toxicities with grade larger or equal to 2. For example, a patient with 5 out of 6 toxicities with grade larger than one was administered a dose reduction, although less than 20%, and a delay of 17 days (see plot for cycle 4). All orange circles, which correspond to four toxicities with grade  $\geq 2$ , lie above the horizontal line

of no delay, which means that there was a delay. Patients with three severe toxicities visualised in light-blue are scattered all over the plots but hardly any of them lies close to standardised dose of one and zero days delay. This means that their therapies were modified accordingly. Yet yellow, blue and red circles could be found everywhere in the plots. This suggests that the reason for therapy modification for these patients is not present in the figure. Maybe unmeasured confounders are present; oncologist may know more about patient's condition than the records could reveal. Another possibility could be that treatment policy varies across hospitals and practitioners. For example, a conservative oncologist could prescribe dose reduction in case of myelosuppression grade 2, while another could prescribe G-CSF infusion (a WBC production stimulating factor) and no dose reduction. The latter hypothesis is generated by a careful inspection of patient records. Indeed, a couple of patients from the control arm received a number of G-CSF infusions with varying doses.

Figure 3.14 illustrates this situation. It visualises treatment trajectory of patients 11 and 1106, and a hypothetical treatment trajectory for patient 1106 who follows the protocol. For each of the scenarios, a time axis is presented. Above it the doses of both drugs are shown as bars which height could vary, presenting dose reductions. Below the time axis, for each cycle the six treatment side effects are listed. Their corresponding severity is represented with bars of different height. A bar of height zero, or a white background, stands for lack of toxicity, i.e. grade 0, or no dose given. The two patients were chosen to have no delays in order not to overload the figure with information, although delays could be represented via shifts along the time axis.

Patient 11 was administered six chemotherapy cycles; surgery was performed on time between cycles 2 and 3. The patient experienced no toxicities throughout cycles one to three, and was given full doses. As a result of cycle 4, the patient develops oral toxicity of grade 1. The treatment was not modified. However, throughout cycle 5 the patient develops grade 1 myelosuppression and grade 2 oral toxicity. Because of the latter, patient 11 is administered 20% dose reduction in cycle 6. The record of this patient is an example of protocol adherence. Yet, not every oncologist would apply dose reduction in case of oral toxicity grade 2. On one hand, because the protocol prescribes dose reduction in case of oral toxicity more severe than grade 2, and on the other because the patient had no side effects from previous cycles and the sixth cycle will be patient's last cycle, i.e. there is no possibility for accumulating toxicities.

On the contrary, patient 1106 is an example of poor adherence to the protocol. This patient experiences toxicities that could justify dose reductions in a number of subsequent cycles. Nevertheless, the patient is always exposed to full dose. In order to explain what could have been done, patient 1106 toxicities record is shown twice on the third time axis. The myelosuppression and infection after cycle 2, both with CTCAE grade equal to 2, should result in dose reduction for cycle 3. Then the subsequent occurrences of these side effects should cause further 20% dose reduction. Finally, as the protocol prescribes,

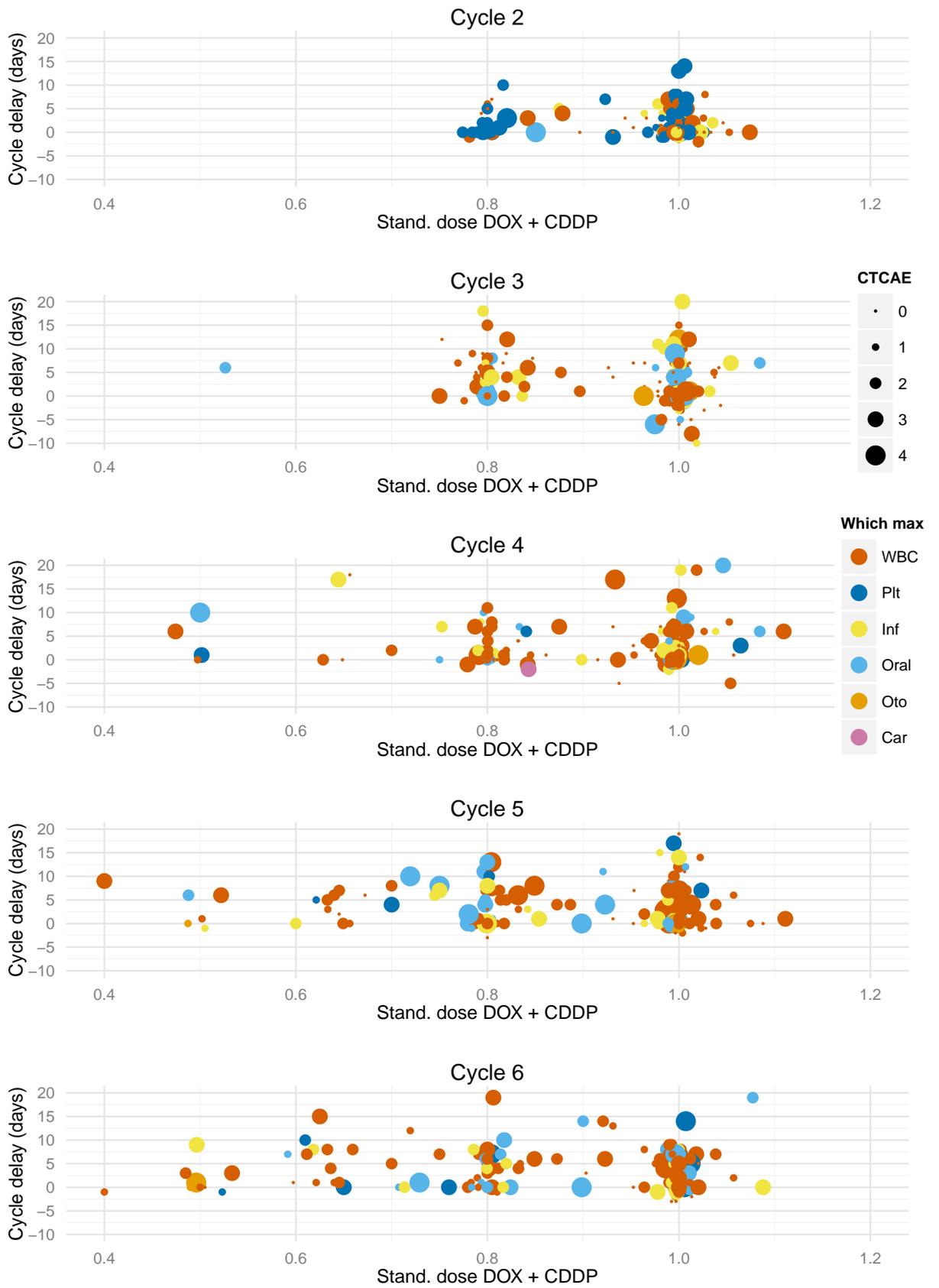


Figure 3.12: Therapy modifications per cycle as severest toxicity from previous cycle

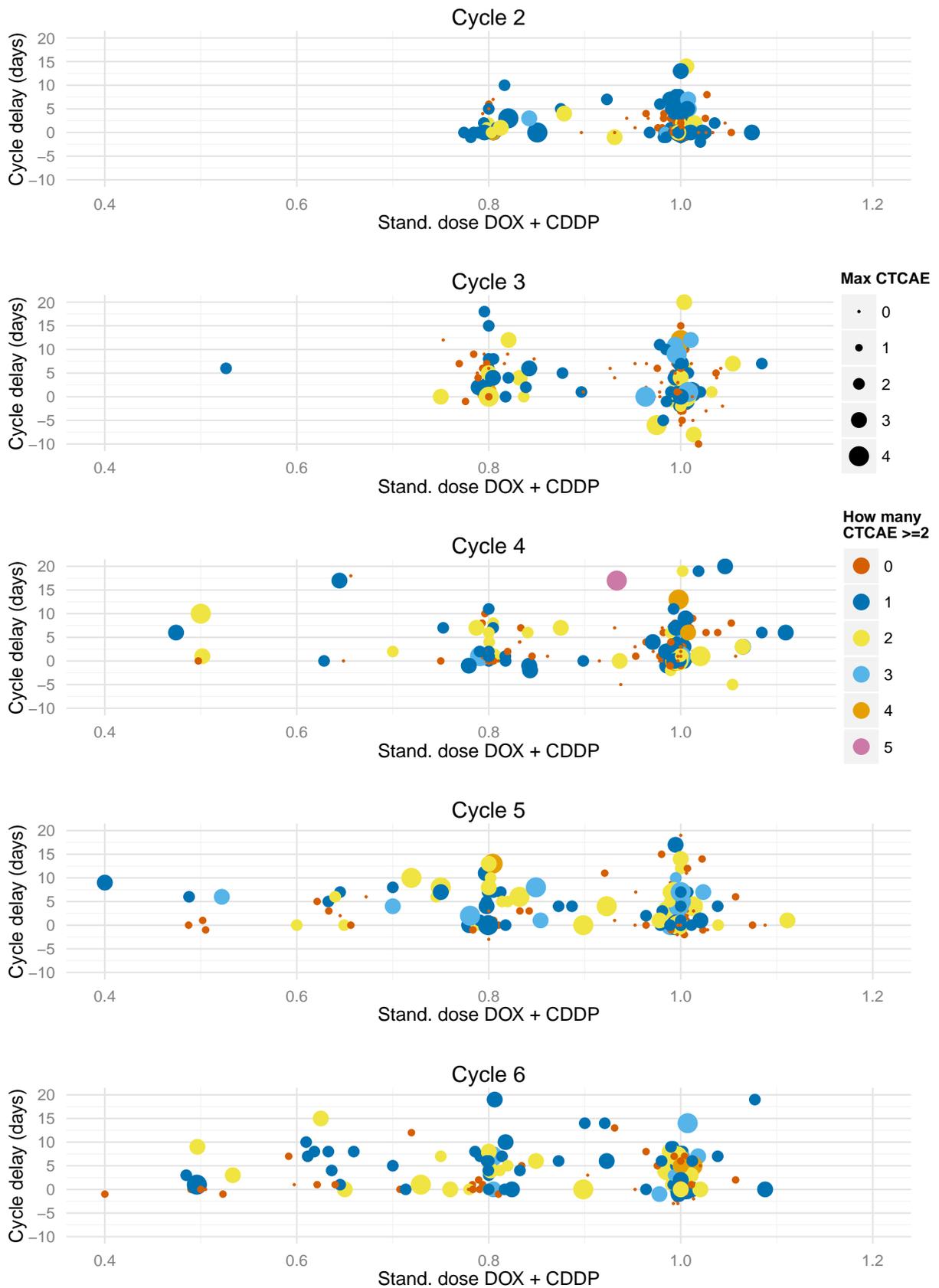


Figure 3.13: Therapy modifications per cycle as number of toxicities from previous cycle with grade larger than one

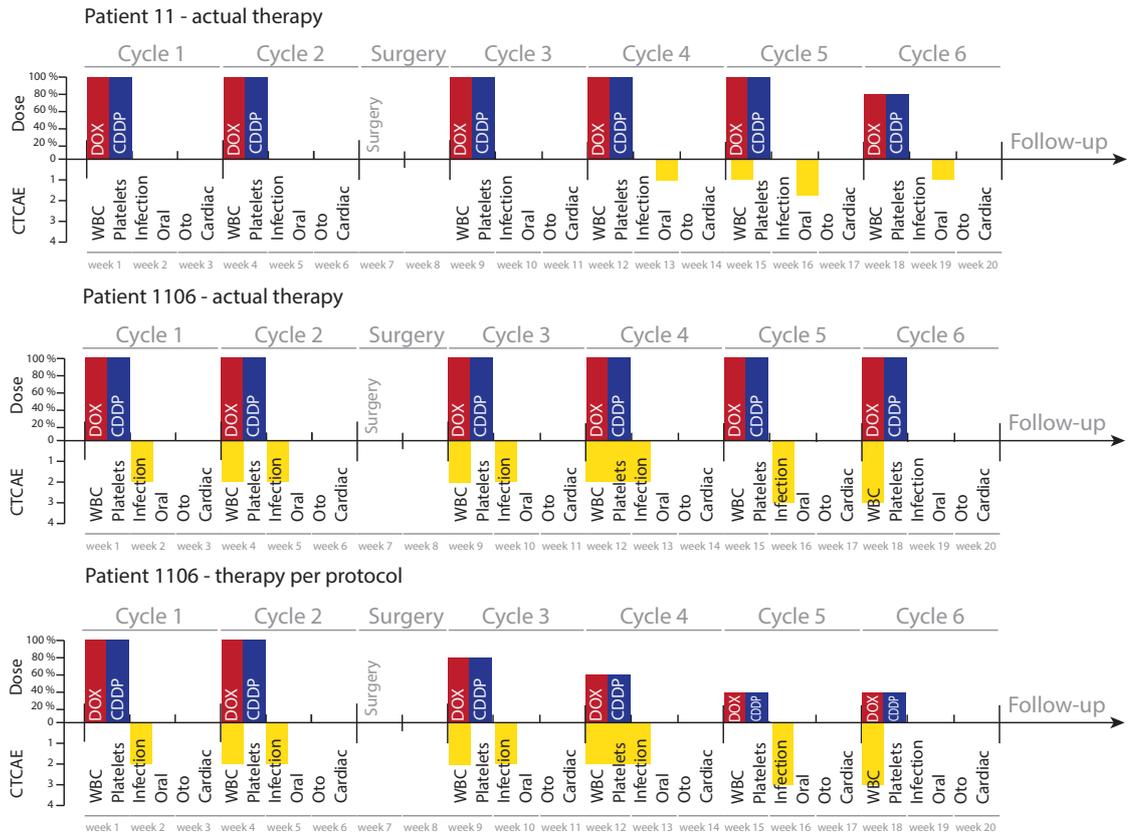


Figure 3.14: Treatment trajectories of patients 11 and 1106, and therapy of patient 1106 according to the protocol

once applied, dose reductions should be preserved at following cycles.

Records do not always agree with researchers' expectations. However, the work done for this thesis gives us opportunity to stress how important field experts' opinion in building models for causal inference is. Still, our believe in the correctness of the presented reasoning, and the records that support our understanding of the process of treatment modification in chemotherapy, permit us the possibility to try and build causal models based on these data.

### 3.5 Missing values, data preprocessing, patients' eligibility

This section is devoted to some technical problems we encountered while working with the data. Sub-section 3.5.1 elaborates on missing values and how we cope with them. Finally, some misleading fields are pointed out in order to show that clinical records should not be taken for granted, and cross-checked wherever possible.

### 3.5.1 Missing values

Follow-up was available for every patient, however HRe is missing for 17% of the patients. In one of the analyses performed in this thesis, HRe is used as an outcome variable and as such is not imputed. This issue forced us to exclude the records of 37 patients from the 218 who meet the trial eligibility criteria.

We had to derive the exposures of interest but still missing values were present. For the computation of delays we need the date of surgery. This is not available for 7.3% of the patients. For some of them an approximation of date of surgery could be retrieved from resected specimen evaluation form. However, this would introduce uncertainty in exposure values. Furthermore, 75% of the patients with missing date of surgery were missing HRe. Most probably these patients did not undergo surgery. Because of these circumstances, no approximation was performed.

For the calculation of standardised dose we need the dose administered in  $mg$  per  $m^2$  of body surface area. Chemotherapy form contains doses in  $mg$  and the following fields: body surface area ( $m^2$ ), weight ( $kg$ ), and height ( $cm$ ). There are no missing doses but some body surface area fields needed to turn administered dose in  $mg/m^2$  were blank. Body surface area is a function of height and weight, and clinicians use a nomogram to compute it. However, the latter is not part of the protocol. Often height and/or weight was available for those patients and combining them with patients' records from previous cycles and the given dose in  $mg$ , we were able to reconstruct their body surface area.

Figure 3.7 shows that some certain toxicities were not graded. After extensive discussion with an oncologist, we have decided to treat those missing values as 0-grades.

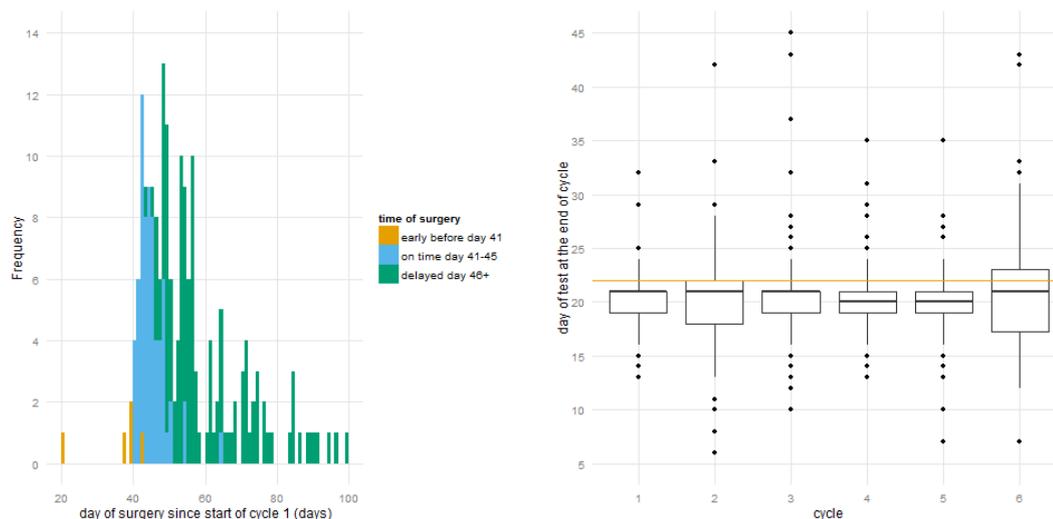
### 3.5.2 Data preprocessing

Some of the available records were misleading. For example, follow-up form contains a field that indicates whether surgery was performed. When contrasted with availability of date of surgery, four patients were indicated to have undergone surgery while not having a record of date of surgery.

An interesting and very important field of surgery form addresses surgery delay. In this field time of surgery could be indicated as: *early* (before day 41), *on time* (between day 41 and day 45), and *delayed* (later than day 45). On one hand, this information an idea about what is considered a delay. A surgery on day 46 or later is equivalent to a delay of 3 or more days. On the other hand, as Figure 3.15a shows, this field could be misleading. As indicated by the colour, there are surgeries performed after day 50 since start of therapy and are still reported as being performed *on time*. Maybe a delay was introduced, and an additional chemotherapy cycle was given. Then surgery was performed on time relative to the end of the additional cycle. However, the benchmarking days are listed in the form and it looks like the reference point for surgery timing is the start of the therapy, and not the start of the second to last cycle.

Another confusing field concerns date of laboratory measurements at the end of each cycle. Chemotherapy form and the protocol state that WBC count, neutrophils, and platelets should be measured on day 22 of each cycle. Nevertheless, this is hardly the

case. The orange horizontal line on Figure 3.15b indicates the 22nd day since start of cycle. The distribution of the day of laboratory test since start of the cycle is highly variable. Often this test is performed earlier than later. In some cases it is recorded before day 10 of the cycle while there is another column in the chemotherapy form which leaves room for test results from day 10. Caution is needed here because these fields are used to evaluate haematological toxicity at the end of each cycle.



(a) Day of surgery coloured by time of surgery (b) Distribution of time to laboratory measurements at the end of cycle

Figure 3.15: Plots of examples of records that should be treated with care

### 3.5.3 Patients' eligibility for analysis

In this section the selection of the sample of patients used in the analyses is described. This is visualised in the flow-chart in Figure 3.16. The study enrolled 497 patients, and randomised 245 to the control arm. The latter is the set used in this thesis. We could not analyse the records of 7 patients because they had no treatment data. Also in the control arm, 19 patients were given G-CSF at least during one cycle. The doses varied, therefore it was hard to correct for it. We decided not to analyse these patients. As discussed in Section 3.3, we need the date of surgery in order to calculate the number of preoperative cycles, cycle durations, and delays. Since we are interested in the effect of delays on different outcomes, we could not use the data from 17 patients with unknown surgery date. As a result, the eligible sample shrunk to 202 patients. For the analyses of HRe we additionally excluded 25 patients without HRe record. For the survival analyses we did not use the data from 9 patients who prematurely terminated treatment due to disease progression. As a result, there are 193 patients, who are eligible for survival analysis.

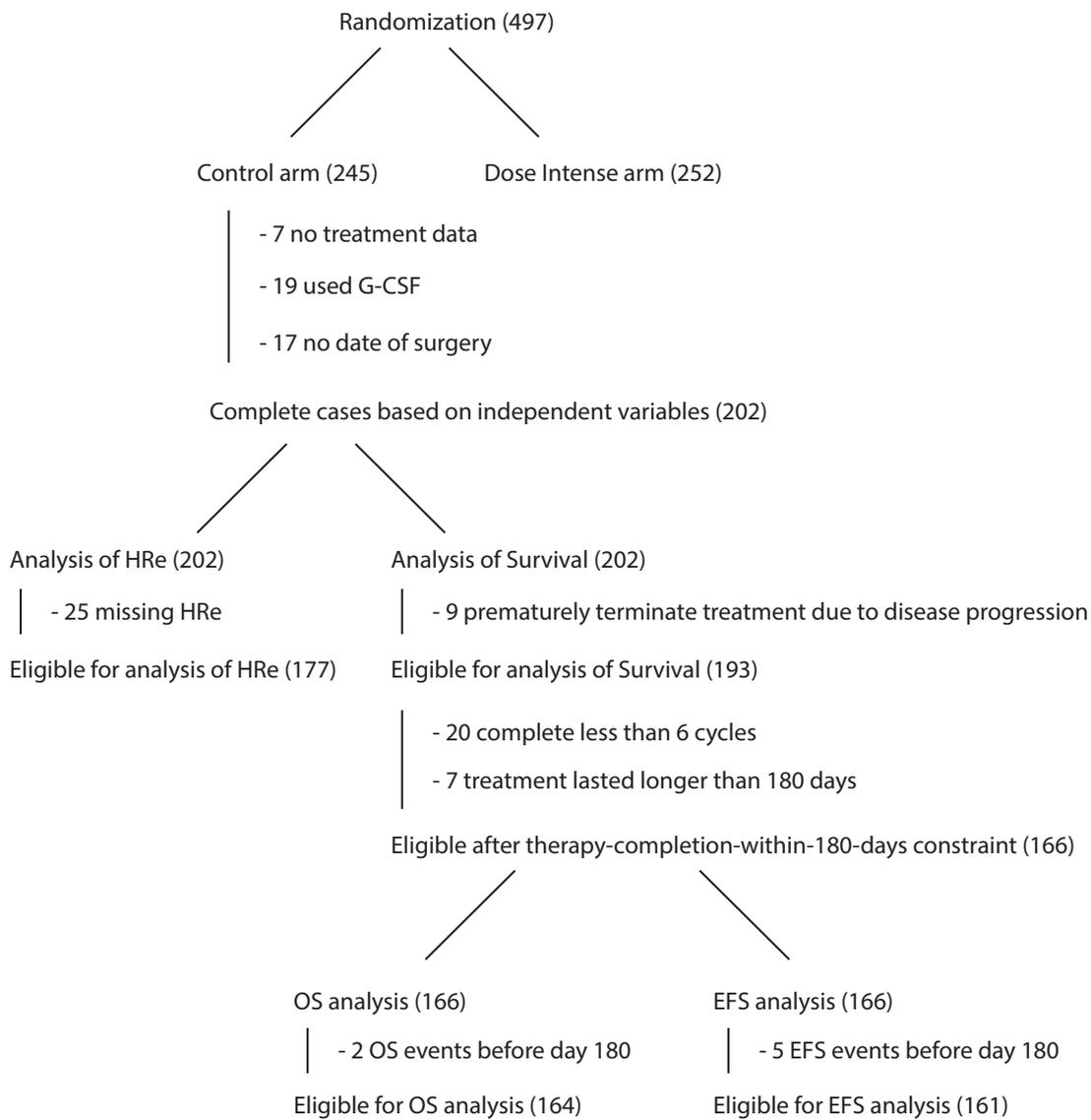


Figure 3.16: Flow-chart of patients selected for analyses

As stated in Chapter 5, we perform survival analysis conditional on therapy completion within 180 days, i.e. landmark analysis. Because of premature treatment termination (for reasons other than disease progression, see Figure 3.4b) we had to exclude 20 patients. The landmark at 180 days further reduces the available sample by 7 patients, whose treatment lasted longer than 180 days. As a result there are only 166 patients left. Further, we distinguish analysis of OS from analysis of EFS. The former is based on 164 patients when 2 patients are excluded since they died before the landmark. The

EFS is estimated by further excluding 5 patients since they experienced an event before the landmark. The analyses of HRe (if without any other constraints) are based on a sample of 177 patients, while for OS and EFS we use the records of 164 and 161 patients, respectively.

## Chapter 4

# MSMs for Histological Response

The present chapter describes how to build MSMs for causal effects on a binary outcome using the dataset described in the previous chapter. Three models are presented in order of increasing complexity. Models' intricacy grows in two directions - number of exposures, and specifics of the data that are taken into account. The presentation of each model starts with an introduction to the problem that the model answers. Then a MSM is developed and estimated. The results are contrasted with those from a model for association.

### 4.1 MSM for causal effect of cycle delays on histological response – a simplified example

This section presents a simple MSM for estimation of the causal effect of chemotherapy cycle delays on histological response (HRe) – a binary variable addressing tumour response to chemotherapy (see Definition 3.1). The model is simple because we restrict the analysis to the subgroup of patients who underwent exactly two preoperative chemotherapy courses, and thus observe only one instance of exposure, namely delay of cycle 2. Cytotoxic drugs given during chemotherapy produce side effects that affect HRe. The study protocol prescribes a cycle delay of one week (or less in case of full recovery) when WBC or platelets count correspond to CTCAE grade 2 or larger, and a delay until full recovery if oral toxicity occurs. These therapy modifications are expected to decrease the severity of the side effects and prevent death due to treatment. Therefore, toxicities are both (1) predictors of subsequent therapy modification, and (2) risk factors for the outcome of interest. The interdependences between cycle delays, toxicities and the outcome are visualised through the DAG on Figure 4.1.

By definition, the first cycle cannot be delayed because its beginning determines therapy time origin. A set of *baseline confounders* determine patient's adverse-effects susceptibility profile. According to expert knowledge, these are patient's age and gender. We denote them by the vector  $\mathbf{V} = (\mathbb{1}_{\{\text{female}\}}, \text{age})$ , where the first element of the vector is a gender indicator function, and the second is a continuous variable represent-

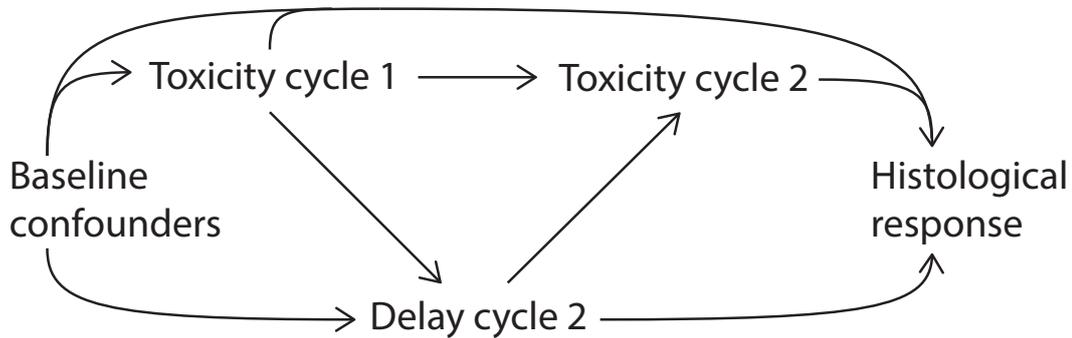


Figure 4.1: DAG describing the causal relationships between cycle delays, confounders and histological response

ing patient’s age. These patient’s characteristics influence how severe side effects the patient will develop; how long it will take him/her to recover from these side effects; and how responsive to chemotherapy the tumour will be, i.e. his/her HRe. This implies that in the DAG on Figure 4.1 edges connect *baseline confounders* with *toxicity cycle 1*, *delay cycle 2*, and *histological response*. Throughout the first cycle, patient’s reaction to chemotherapy is measured by a set of toxicities. As listed in Chapter 3, side effects considered most important by field experts are myelosuppression, infection, oral and cardiac toxicity, and ototoxicity. The severity of each one is scored as a CTCAE grade. Based on these toxicities the consecutive cycle is delayed or not. However, if a patient experienced a severe toxic side-effect as a result of the first chemotherapy cycle, then he/she is likely to experience it again throughout cycle 2, see for example Figure 3.14. A potential explanation could be that patients probably need more time to recover from the developed toxicities than the protocol allows, and thus patients are susceptible to accumulating toxicities over time. On the contrary, low or no toxicity indicate tolerance to chemotherapy. Correlation between toxicities from consecutive cycles is presented in Figure 4.1 via an edge between vertices *toxicity cycle 1* and *toxicity cycle 2*. In some patients, toxicities might indicate therapy effectiveness. This is because of the well-known clinical saying “If the body suffers then the tumour suffers too”. Chemotherapy targets the same apoptotic pathways of tumourous and normal cells. Apoptosis of cancer cells is the goal of the therapy, while apoptosis of non-tumorous cells gives rise to a number of toxicities (mainly bone marrow suppression). Since toxicities affect the outcome, edges between toxicities and *histological response* are drawn. Likewise, if a delay is applied, the body has more time to recover but the tumour might regrow. As a result, both the severity of subsequent toxicity is decreased and Poor (histological) Response (PR) is more likely to be observed. These two relations are expressed through edges linking *delay cycle 2* to *toxicity cycle 2* and *histological response*.

### 4.1.1 Investigation of time-dependent confounding and exposure – confounder feedback

The causal pathways in Figure 4.1 could be used to build a MSM. Chapter 2 discusses the use of MSMs in case of time-dependent confounding and exposure-confounder feedback. Before diving into model building and estimation we should investigate whether these phenomena are present. We investigate whether toxicities are risk factors for HRe, whether they influence administration of delays, and whether delays are associated with less severe toxicities from the subsequent chemotherapy courses. Building association models for that purpose is enough [2].

In order to determine whether treatment side effects predict HRe, we build a logistic regression model. In this model the dependent variable is HRe and the independent ones are toxicities from both cycles, baseline confounders, and delay of cycle 2. We correct the effects of toxicities for the baseline covariates and delays in order to block the indirect effect pathways that go through these nodes in the DAG. Before formulating this model in a mathematical way, we introduce some notation.

**Definition 4.1. Outcome**  $Y$  is a dichotomous variable denoting the outcome, histological response, i.e.

$$Y = \mathbb{1}_{\{\text{good histological response}\}} \begin{cases} 1 & , \text{ if } \leq 10\% \text{ viable tumour,} \\ 0 & , \text{ if } > 10\% \text{ viable tumour,} \end{cases} \quad (4.1)$$

where  $\mathbb{1}_{\{\cdot\}}$  is the indicator function.

The study protocol prescribes cycle delays if haematological toxicity grade 2 or higher occurs. Haematological toxicity, also called myelosuppression, is expressed in toxicity grades for leucopenia (low WBC count) and thrombocytopenia (low platelets count). In the sample of patients with surgery after two chemotherapy courses, i.e. 135 patients, only 2 experience platelets shortage as a result of cycle 1 and are assigned a thrombocytopenia CTCAE grade equal to 3. This suggests that cycle-two delays due to thrombocytopenia are a negligible fraction. Nevertheless, myelosuppression could be addressed by modelling WBC count.

**Definition 4.2. Leucopenia**  $L_k^1$  is a dichotomous variable denoting occurrence of leucopenia throughout cycle  $k$ , i.e.

$$L_k^1 = \mathbb{1}_{\{\text{leucopenia in cycle } k\}} \begin{cases} 1 & , \text{ if WBC count CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & , \text{ if WBC count CTCAE grade } < 2 \text{ for cycle } k. \end{cases} \quad (4.2)$$

Additionally, a cycle delay is recommended if oral toxicity, also called mucositis, occurs. The protocol does not indicate which CTCAE grade implies cycle delay. Field experts advocate that action is needed only when a CTCAE grade larger than two is recorded. However, due to lack of variability in the sample at hand, we had to make a compromise and consider grades larger than one as indicators of oral toxicity. In fact, estimating the effect of oral toxicity of grade larger than two will not be possible with these data, since such a severe toxicity occurs only in six patient-cycles.

**Definition 4.3. Oral toxicity**  $L_k^2$  is a dichotomous variable denoting occurrence of oral toxicity throughout cycle  $k$ , i.e.

$$L_k^2 = \mathbb{1}_{\{\text{mucositis in cycle } k\}} \begin{cases} 1 & , \text{ if oral toxicity CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & , \text{ if oral toxicity CTCAE grade } < 2 \text{ for cycle } k. \end{cases} \quad (4.3)$$

In Chapter 3 six types of treatment side effects that are considered most important by oncologists are described. However, those that are left are either not related to administration of delays (infection), or are late toxicities, i.e. occur in later cycles (cardiac and ototoxicity).

For simplicity of exposition, we can combine all treatment side effects into a vector of confounders, i.e  $\mathbf{L}_k = (L_k^1, L_k^2)$ , for  $k = 1, 2$ . Further, let us denote the whole confounder history up to cycle 2 by  $\bar{\mathbf{L}}_2$ . That is  $\bar{\mathbf{L}}_2 = (L_1^1, L_1^2, L_2^1, L_2^2)$ .

The exposure of interest in this model is cycle-two delay. This definition is motivated by our desire for consistency with previous published analyses of this dataset [28], although expert knowledge suggests that short delays of 3 days are unlikely to have an effect on the outcome or on subsequent toxicities. Although, Definition 4.4 treats equally long and short delays, the majority of the delays are short (see Figure 3.9a). Therefore, we claim to make inference about short delays.

**Definition 4.4. Cycle delay**  $A$  is a dichotomous variable denoting delay of cycle 2, i.e.

$$A = \mathbb{1}_{\{\text{cycle delay } \geq 3 \text{ days}\}} \begin{cases} 1 & , \text{ if it starts later than day 23 of the first cycle} \\ 0 & , \text{ if it starts before day 24 of the first cycle.} \end{cases} \quad (4.4)$$

Based upon the definitions listed above, we can build a model that quantifies the association between toxicities and histological response. HRe is a binary variable, hence we use a logistic regression model

$$P(Y = 1 \mid \bar{\mathbf{L}}_2, A, \mathbf{V}) = \text{expit}(\delta_0 + \boldsymbol{\delta}_1^\top \bar{\mathbf{L}}_2 + \delta_2 A + \delta_3 \mathbb{1}_{\{\text{female}\}} + \delta_4 \text{age}), \quad (4.5)$$

where  $\text{expit}(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}$ .

In Model (4.5) age is treated as a continuous variable centred at 16 years to enable interpretation of the model intercept. Parameter estimates are displayed in Table 4.1. Presented numbers correspond to  $\delta$ -parameters, i.e. log-odds of good HRe versus poor HRe for a unit increase in covariate's value. Occurrence of myelosuppression throughout any cycle decreases the probability of good HRe. The same applies for mucositis. The standard error of all effect estimates are too large to allow us to reject the null hypothesis of no association between toxicities and the outcome. Besides leucopenia in cycle 1, all toxicities have clearly non-zero effect on HRe. This suggests that the inferential weakness observed in Table 4.1 might be grounded on the lack of power of the statistical tests. Moreover, only 42 patients achieve good HRe and this means that the general rule of 10 events per variable in logistic regression Model (4.5) is violated.

Nevertheless, Model (4.5) serves as an ad-hoc approach to evaluate to what extent the data at hand supports the solid clinical understanding of chemotherapy personalisation for osteosarcoma patients.

Table 4.1: Toxicities effect estimates on histological response

	Estimate*	SE(Estimate)	<i>p</i> -value
(Intercept)	-0.424	0.304	0.163
Leucopenia in cycle 1	-0.065	0.622	0.917
Leucopenia in cycle 2	-0.626	0.493	0.204
Mucositis in cycle 1	-0.550	0.695	0.429
Mucositis in cycle 2	-0.919	0.692	0.184
Delay of cycle 2	0.511	0.458	0.265
Female	-0.548	0.417	0.189
Age	-0.064	0.035	0.072

\* log-odds

As a subsequent step we build a logistic regression model for allocation of cycle delays to check whether this decision is based on toxicities. In this model we have to correct for the baseline covariates in order to block the potential back-door from cycle-one toxicities to delays through the baseline confounders. That is, we fit model

$$P(A = 1 \mid \mathbf{L}_1, \mathbf{V}) = \text{expit}(\delta_0 + \boldsymbol{\delta}_1^\top \mathbf{L}_1 + \delta_2 \mathbb{1}_{\{\text{female}\}} + \delta_3 \text{age}). \quad (4.6)$$

Note that this model contains only the side effects of treatment throughout cycle 1. If we include toxicities from cycle 2, we will condition on the future and that would be wrong.

Parameter estimates of Model (4.6) are presented in Table 4.2. As expected, leucopenia increases the probability of delay allocation. In contrast, mucositis is associated with lack of delays, i.e. the association parameter estimate is negative. This is probably driven by the low cut-off point for dichotomisation of oral toxicity. In other words, in the sample patients with  $L_1^2 = 1$  have an oral toxicity CTCAE grade equal to 2 and were not administered a delay, thus the association parameter estimate is negative. Yet, none of the effect estimates is significantly different from zero.

Table 4.2: Toxicities effect estimates on allocation of cycle delays

	Estimate*	SE(Estimate)	<i>p</i> -value
(Intercept)	-1.290	0.292	0.000
Leucopenia in cycle 1	0.739	0.559	0.187
Mucositis in cycle 1	-0.363	0.687	0.597
Female	-0.209	0.444	0.638
Age	-0.009	0.035	0.802

\* log-odds

Since toxicities are also modelled as binary variables, we use logistic regression to quantify the effect of cycle-two delay on occurrence of a subsequent treatment side effect. In this example there are two toxicities, hence we have to build two models. In order to block all possible indirect effect pathways we need to correct the effect of the exposure for the baseline covariates and the toxicities in cycle 1. This is done by using the generic model

$$P(L_2^j = 1 \mid A, \mathbf{L}_1, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 A + \delta_2^\top \mathbf{L}_1 + \delta_3 \mathbb{1}_{\{\text{female}\}} + \delta_4 \text{age}), \quad (4.7)$$

for  $j = 1, 2$ , i.e. we use the same set of independent variables to model the occurrence of leucopenia and oral toxicity, respectively.

For presentation purposes, in Table 4.3 we display only the exposure effect estimates obtained from fitting the two models in Equation (4.7). These data do not provide evidence that cycle-two delay is associated with subsequent occurrence of leucopenia. On the contrary, allocation of delay before the start of cycle 2 decreases the probability of mucositis in cycle 2. However, this effect estimate is not statistically significantly different from zero.

Table 4.3: Cycle-two delay effect estimates on occurrence of severe toxicities

Effect of cycle-two delay on:	Estimate*	SE(Estimate)	<i>p</i> -value
Leucopenia in cycle 2	0.023	0.503	0.963
Mucositis in cycle 2	-0.182	0.695	0.794

\* log-odds

The analyses in this section provide evidence for most of the relationships that are known to exist between toxicities, cycle delays, and HRe. We recognise the lack of statistical power due to the small sample size, and the limitations that these analyses possess due to imprecise phenomenon definitions driven by the low variability. However, these analyses serve mainly as exploratory technique, and will later be used to enhance the plausibility of correct model specification assumption that MSMs rely on. In the next section we use some of the models presented here to estimate IPTWs and build a MSM for the effect of cycle-two delay on HRe.

#### 4.1.2 Model formulation and estimation

MSMs are models for the marginal distribution of counter-factual outcomes. As elaborated on in Chapter 2, counter-factual outcomes are the outcomes that we would have observed had the patient followed a certain treatment trajectory.

**Definition 4.5. Counter-factual outcome**  $Y_a$  is a dichotomous variable denoting the outcome, histological response, arising from exposure  $A = a$ .

$$Y_a = \mathbb{1}_{\{\text{good histological response, arising from exposure } A = a\}} \begin{cases} 1 & , \text{ if } \leq 10\% \text{ viable tumour,} \\ 0 & , \text{ if } > 10\% \text{ viable tumour,} \end{cases} \quad (4.8)$$

where  $a$  denotes the realisation of the random variable  $A$ , taking values 0 or 1.

Definition 4.5 implies that each patient has two counter-factual outcomes - one if patient's second cycle was delayed, denoted by  $Y_1$ , and one if patient's second cycle was not delayed, denoted by  $Y_0$ .

With this notation a marginal structural logistic model for the causal effect of cycle-two delay on HRe is defined as follows:

$$P(Y_a = 1) = \text{expit}(\beta_0 + \beta_1 A). \quad (4.9)$$

Model (4.9) is estimated by fitting a weighted regression model for the factual outcome, denoted by  $Y$ , according to Definition 4.1. That is, estimating the parameters of model

$$P(Y = 1 \mid A = a) = \text{expit}(\gamma_0 + \gamma_1 A) \quad (4.10)$$

in a pseudo-population created by IPTWs.

These weights are estimated by building a model for allocation of the exposure, delay of cycle 2, given all time-dependent and baseline confounders. The model for exposure allocation is used to estimate the probability of the observed exposure. Its inverse, i.e.  $1/\text{the probability}$ , is equal to the pursued weight. As explained in Section 2.2.1, this weighting is designed to eliminate the confounding, since inverse-probability-of-treatment weighting mimics randomisation of the exposure with respect to the measured confounders. This means that all spurious associations attributable to the original sample are not present in the pseudo-population, and the association parameter between the exposure and the outcome can be interpreted in a causal way.

In the current example the model for exposure allocation is the same as Model (4.6). However, for the estimation of the weights we need to pre-process the results from Model (4.6), used to quantify the effect of the toxicities on cycle-two delays. This model estimates the conditional probability of delay, i.e.  $P(A = 1 \mid \cdot)$ , while the weights are comprised of probabilities of observed exposure. As we have seen, not everybody's second cycle was delayed. This means that the probability of the observed exposure for some patients is the probability of delay, and for the rest - the probability of lack of delay. Let us denote the probability of cycle-two delay estimated through Model (4.6) by  $p_{a=1}$ . Then, by introducing a patient index  $i$ , the subject-specific weight is equal to

$$W_i = \frac{1}{p_{A_i=a_i}} = \frac{1}{[(p_{i, a=1})^{(A_i)}] \times [(1 - p_{i, a=1})^{(1-A_i)}]}, \quad (4.11)$$

where the denominator of the fraction resembles binomial likelihood and is equal to  $P(A_i = 1 \mid a_i = 1)$  for patients with delayed cycle, or to  $P(A_i = 0 \mid a_i = 0)$  for patients whose second cycle was not delayed. In Formula (4.11),  $p_{A_i=a_i}$  denotes the the probability of the observed exposure.

Table 4.4 summarises the distribution of the estimated quantities that comprise the weights. The model for exposure allocation is not able to distinguish patients with delay from those without since all estimated probabilities of delay are very low. This means

Table 4.4: Distribution of estimated probabilities of cycle-two delay, probabilities of observed exposure, and final weights

	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
$\hat{p}_{a=1}$	0.117	0.185	0.210	0.215	0.222	0.374
$\hat{p}_{A_i=a_i}$	0.117	0.674	0.784	0.669	0.813	0.876
$\hat{W}_i$	1.141	1.230	1.275	2.007	1.483	8.523

that the selected toxicities from cycle 1, and the baseline confounders do not explain the variability of  $A$ , although cycle-two delay is twice more likely in case of leucopenia throughout cycle 1 ( $OR = e^{0.739} = 2.09$ ). Most patients' second cycle was not delayed and the low probabilities of delay are translated to high probabilities of observed exposure. The top 75% of  $\hat{p}_{A_i=a_i}$  range between 0.674 and 0.876. The patient with the highest probability of observed exposure receives the lowest weight ( $1/0.876 = 1.141$ ), while the patient with the lowest probability of observed exposure receives the highest weight ( $1/0.117 = 8.523$ ). High values of the probabilities of observed exposure correspond to small weights. This is a consequence of the functional form of the weights.

The second cycle of the patient with the lowest estimated probability of cycle-two delay was delayed ( $\min(\hat{p}_{a=1}) = \min(\hat{p}_{A_i=a_i})$ ). This patient's record is a counter-example of the predominant relationship between the toxicities and delay allocation (see Table 4.2). If a patient's record is over-expressed, it may counter-balance and neutralise the observed association between toxicities and cycle delays. As a result, the technique uses 8.523 copies of a patient's record to form the pseudo-population in which there is no confounding.

### 4.1.3 Results and discussion

The causal effect of cycle-two delay on HRe is estimated by fitting weighted logistic regression Model (4.10). It is instructive to compare it with the association parameter estimated by fitting the same Model (4.10) with subject-specific weights all equal to one, i.e. through unweighed regression. The two odds-ratio (OR) estimates are shown in Table 4.5. Robust standard error estimates, obtained using the 'sandwich' variance-covariance estimator, are used to construct 95% confidence intervals (CI). The delay of the second cycle is estimated to cause on average 1.824 times more often a GR than a cycle on time causes. The OR for association is similar. This might suggest that the effect of cycle delay might not be confounded with the effect of toxicities from cycle 1 on HRe. This is in line with the results in Section 4.1.1, where we quantify the extent of association between the toxicities and the delays of the subsequent cycle. Furthermore, these results are not in line with our understanding of the effect of large cycle delays on HRe. Delays should allow the body to recover but also the tumour to regrow. Therefore, delays should have a negative effect on HRe. However, we recognise that all but 7 of the 29 delayed cycles were delayed for no more than 5 days (see Figure 3.9a). Expert knowledge suggests that longer delays are needed to allow the tumour to regrow and

increase the probability of a PR. Nevertheless, the confidence intervals of both odds-ratios are so wide that we cannot reject the hypothesis of no direct effect.

Table 4.5: Causal and association odds-ratio estimates for the effect of cycle-two delay on histological response

	OR	95% CI	<i>p</i> -value
Causal	1.824	[0.761 : 4.372]	0.180
Association	1.788	[0.761 : 4.203]	0.185

As Figure 3.5a shows, patients' preoperative treatment trajectories have different length. Some were operated prematurely, and others receive more than two chemotherapy cycles before surgery. The analyses in this section use only patients with exactly two preoperative courses. The results from such an analysis might be invalidated by potential selection bias. The model in the next section incorporates length of treatment trajectory into the analysis. Furthermore, it is a model for joint causal effect of cycle delays and administration of dose reductions/discontinuations.

## 4.2 MSM for joint causal effect of cycle delays and dose reductions on histological response

The variable length of patients' treatment trajectories is very difficult to deal with. This problem was eliminated for simplicity of the analysis in the previous section by using only patients with exactly two preoperative chemotherapy cycles. This subgroup is the majority of patients with available date of surgery and HRe.

The distribution of the number of preoperative cycles is listed in Table 4.6. Patients with 1 or 6 preoperative cycles (3 patients in total) may be considered a negligible proportion, yet there are 39 patients with 3 or 4 preoperative cycles which would better be included in the analysis.

Table 4.6: Distribution of number of preoperative cycles

Number of preoperative cycles	1	2	3	4	5	6
Frequency	2	135	30	9	0	1

Varying length of preoperative treatment trajectory introduces two issues. First, exposure becomes longitudinal, i.e. it may occur multiple times throughout treatment. This poses the question - how to model cumulative exposure? In the context of the current example, the answer is relatively easy. If we sum the number of occurrences of a given exposure, cumulative exposure becomes a continuous variable taking discrete values, e.g. from zero to three. The regression coefficient for this variable is then interpreted as the effect of an additional instance of exposure administration. Second, varying length of preoperative treatment trajectory makes us wonder what causes administration of additional preoperative courses of chemotherapy. The study protocol addresses the

possibility of surgery delay mainly in the context of the dose-intense arm (see Section 3.1).

It is suggested that surgery might be delayed for patients who cannot recover from myelosuppression on time, or any other reason merits a week's delay to allow recovery from the treatment side-effects. Furthermore, it is explicitly stated that an additional preoperative course may be given to patients with significantly delayed surgery due to administrative reasons. Indeed, administrative reason for cycle delay was registered for 17 out of the 40 patients with more than two preoperative cycles. For the remaining 23 patients other reasons are listed, among which haematological toxicity and other not-toxicity-related reason are most frequent. The protocol also states that surgery could be performed earlier in case of progressing disease under chemotherapy. Tumour progression as reason for early/delayed surgery is recorded for three patients out of the 177 with available date of surgery and HRe. Two of these were operated after the first chemotherapy cycle. At the same time, six patients in total terminate treatment prematurely because of disease progression. In conclusion, extension of preoperative treatment is both (i) predicted by haematological toxicity, and (ii) a risk factor for HRe [28], therefore we have confounding. Administrative issues are still the most common cause of surgery delay after an additional chemotherapy course, and could be thought of as appearing at random conditional on patient's characteristics. Yet, it is not clear whether this reason refers to surgery delay in terms of days since anticipated surgery day (i.e. what we will call surgery delay in Section 4.3), or in terms of number of preoperative cycles. In spite of this fact, there is some evidence that prolonged preoperative treatment improves HRe [28].

The construction of the MSM that follows relies upon the causal diagram presented in Figure 4.2. As in the previous model, *toxicity* from cycle  $k$  predicts *exposure* in cycle  $k + 1$ , and the *exposure* in cycle  $k + 1$  predicts the *toxicity* during cycle  $k + 1$ . In this example first a preoperative cycle should be administered in order to be delayed or having reduced dose. With this statement we try to address the difference in the number of preoperative cycles. First, we excluded the two patients who were operated after a single course of chemotherapy since they cannot be exposed (the first cycle cannot be delayed and always administers full dose). Every patient in our sample follows two or more cycles before going to surgery. After their first chemotherapy cycle, we observe their exposure during the second cycle and the subsequent toxicities. When the latter are observed, the clinician in charge decides whether the subsequent treatment will be surgery or chemotherapy, i.e. we observe whether surgery is scheduled. This relationship is depicted via an edge between the vertices *toxicity cycle 2* and the first *scheduled surgery* node. If a surgery is scheduled, *scheduled surgery* node is activated, and this blocks the transmission of effects to the right of the *scheduled surgery* node. As a result, we directly observe HRe. However, if surgery is not scheduled, i.e. another chemotherapy course is administered, we observe *exposure cycle 3* and the corresponding toxicity. This relationship is visualised via the edge between the first *scheduled surgery* node and *exposure cycle 3* node. This means that the effect of toxicities on exposure is mediated through a *scheduled surgery* indicator. The process continues in the same

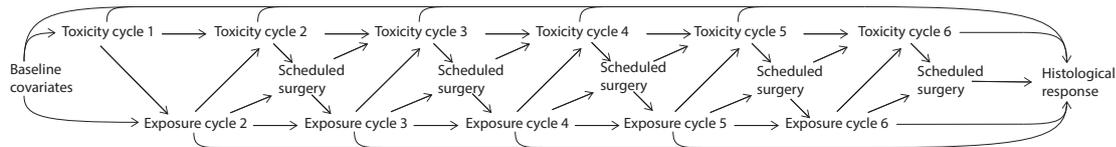


Figure 4.2: DAG describing the causal relationships between exposures, confounders and histological response

manner until the surgery is scheduled and performed. The causal graph in Figure 4.2 represents all possible preoperative treatment scenarios. The baseline covariates determine patient’s predisposition to developing treatment side-effects and thus are connected with each subsequent node on the graph. HRe depends on the whole treatment history. This is graphically depicted by extending a directed edge from every node to *histological response*.

In the present example the aim is to assess the joint causal effect of cycle delays and dose reductions. Analogously to toxicity nodes, exposure nodes incorporate all elements of the exposure – cycle delay and dose reduction. When building a MSM and the corresponding DAG, it is very important to understand the temporal order of the components in the causal pathway. To further clear ideas on how each element appears within a cycle, Figure 4.3 is presented. The central elements in black show the relationship between exposure and toxicity within a cycle. Gray elements are used to place the components of main interest in the context of across-cycles dependence. Dashed lines connect absent items with the present ones. In chemotherapy there is a distinct temporal order of administration of the exposures of interest. First, a cycle starts. Then its date of start indicates whether the cycle was delayed. However, the decision to delay the cycle was taken prior to its start. Expert knowledge and the trial protocol suggest that a cycle is postponed if the patient is too weak to tolerate another course of chemotherapy. Whether this is the case is determined at the end of the previous cycle, i.e. when all side-effects of the previous chemotherapy course are determined. Once a cycle begins and its delay is observed, the dose to be given is determined. This should already be known from patient’s condition at the end of the previous cycle. Yet, at the first day of the new cycle the patient’s condition might have changed and this could further influence the decision on drug dosage (or treatment schedule). Unfortunately, the laboratory measurements taken just prior to the start of a chemotherapy course (usually on the same day or the day before) do not reveal these potential last-minute treatment updates because they are used to justify that a new cycle can begin. As such they hardly ever indicate presence of haematological toxicity. In either case, this proves the temporal order of administration of the two components of the exposure.

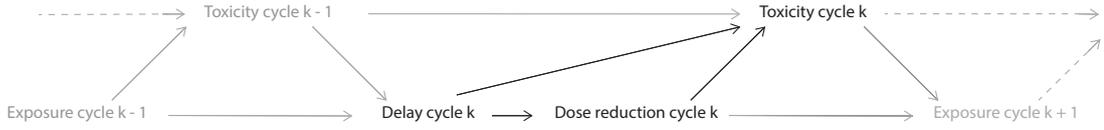


Figure 4.3: Temporal order within a cycle and interrelationships between exposures and confounders

#### 4.2.1 Investigation of time-dependent confounding and exposure – confounder feedback

In this section we build models to evaluate the association between confounders and HRe, and to assess the presence of feedback between confounders and exposures. The association between toxicities and HRe could be quantified with a regression analysis. MSMs are most frequently used in survival analysis. In such a situation, in order to study the relationship between a time-dependent confounder and the event of interest, a pooled logistic regression model might be used (although other methods can be employed). This methodology considers each instance of measurement of the time-dependent covariate as an independent observation. The end-of-study outcome, usually death, could then be traced back to each instance of measurement. In particular, if the patient was treated, he/she must have been alive. Then the outcome indicates "alive" in every instance of observation prior to patient's death. Whereas the same cannot be said for HRe. Because of the possibility for tumour regrowth between chemotherapy cycles, we cannot reject the hypothesis that if a patient was resected one cycle earlier he/she could not have had a GR irrespective of the observed response. The number of viable tumour cells decreases in steps over time, where each drop corresponds to a chemotherapy cycle. In contrast to step functions used to express survival curves where the curve is horizontal between its downward jumps, tumour viability increases between cycles [32]. That is why pooled logistic regression cannot be used to assess the association between a time-dependent covariate and an outcome like HRe.

We should not forget that patients have different length of preoperative treatment. This further complicates our inference. A way out of this situation is to consider cumulative toxicity and correct its effect for the length of treatment trajectory. Although other forms of expressing toxicities might be considered (e.g. mean or maximum through the treatment period), they all take away the cycle-wise nature of toxicity, and thus are considered inferior. Similarly to cumulative exposure, cumulative toxicity could be defined as the number of instances of certain toxicity. A model for toxicities' effects on HRe is a logistic regression with the number of occurrences of each toxicity, the number of occurrences of each exposure, the baseline covariates and the number of preoperative cycles as explanatory variables.

Let us formally define cumulative toxicities in the following way. We denote the number of preoperative cycles by  $s$ .

**Definition 4.6. Cumulative leucopenia**  $\mathcal{L}^1 = \sum_{k=1}^s L_k^1$  is a discrete variable denot-

ing the number of occurrences of leucopenia throughout cycle  $k$ , where  $L_k^1$  denotes

$$L_k^1 = \mathbb{1}_{\{\text{leucopenia in cycle } k\}} \begin{cases} 1 & , \text{ if WBC count CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & , \text{ if WBC count CTCAE grade } < 2 \text{ for cycle } k. \end{cases} \quad (4.12)$$

**Definition 4.7. Cumulative thrombocytopenia**  $\mathcal{L}^2 = \sum_{k=1}^s L_k^2$  is a discrete variable denoting the number of occurrences of thrombocytopenia throughout cycle  $k$ , where  $L_k^2$  denotes

$$L_k^2 = \mathbb{1}_{\{\text{thrombocytopenia in cycle } k\}} \begin{cases} 1 & , \text{ if platelets count CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & , \text{ if platelets count CTCAE grade } < 2 \text{ for cycle } k. \end{cases} \quad (4.13)$$

**Definition 4.8. Cumulative oral toxicity**  $\mathcal{L}^3 = \sum_{k=1}^s L_k^3$  is a discrete variable denoting the number of occurrences of oral toxicity throughout cycle  $k$ , where  $L_k^3$  denotes

$$L_k^3 = \mathbb{1}_{\{\text{mucositis in cycle } k\}} \begin{cases} 1 & , \text{ if oral toxicity CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & , \text{ if oral toxicity CTCAE grade } < 2 \text{ for cycle } k. \end{cases} \quad (4.14)$$

**Definition 4.9. Cumulative infection**  $\mathcal{L}^4 = \sum_{k=1}^s L_k^4$  is a discrete variable denoting the number of occurrences of infection throughout cycle  $k$ , where  $L_k^4$  denotes

$$L_k^4 = \mathbb{1}_{\{\text{infection in cycle } k\}} \begin{cases} 1 & , \text{ if infection CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & , \text{ if infection CTCAE grade } < 2 \text{ for cycle } k. \end{cases} \quad (4.15)$$

**Definition 4.10. Cumulative ototoxicity**  $\mathcal{L}^5 = \sum_{k=1}^s L_k^5$  is a discrete variable denoting the number of occurrences of ototoxicity throughout cycle  $k$ , where  $L_k^5$  denotes

$$L_k^5 = \mathbb{1}_{\{\text{ototoxicity in cycle } k\}} \begin{cases} 1 & , \text{ if ototoxicity CTCAE grade } \geq 1 \text{ for cycle } k \\ 0 & , \text{ if ototoxicity CTCAE grade } = 0 \text{ for cycle } k. \end{cases} \quad (4.16)$$

In the same way cumulative exposures could be defined but summing from cycle 2 until the last preoperative cycle.

**Definition 4.11. Cumulative cycle delay**  $\mathcal{A}^1 = \sum_{k=2}^s A_k^1$  is a discrete variable denoting the number of delayed cycles, i.e.

$$A_k^1 = \mathbb{1}_{\{\text{cycle } k \text{ delay } \geq 3 \text{ days}\}} \begin{cases} 1 & , \text{ if it starts later than day 23 of cycle } k - 1 \\ 0 & , \text{ if it starts before day 24 of cycle } k - 1. \end{cases} \quad (4.17)$$

**Definition 4.12. Cumulative dose reduction**  $\mathcal{A}^2 = \sum_{k=2}^s A_k^2$  is a discrete variable denoting the number of cycles with reduced dose, i.e.

$$A_k^2 = \mathbb{1}_{\{\text{dose reduction in cycle } k\}} \begin{cases} 1 & , \text{ if dose given in cycle } k \leq 86\% \text{ of nominal dose} \\ 0 & , \text{ if dose given in cycle } k > 86\% \text{ of nominal dose} \end{cases} \quad (4.18)$$

Table 4.7 presents the distribution of cumulative cycle delays and cumulative dose reductions. It states that none of the preoperative cycles of 122 patients was delayed, 41 patients had only one delayed cycle, ect., and there are in total 70 delayed cycles. Analogously, 141 patients received full dose during every preoperative cycle, the dose of 30 patients was reduced exactly once, one had three preoperative cycles with reduced dose, and in total 43 dose reductions were administered.

Table 4.7: Distribution of cumulative exposure

	0	1	2	3	Total cycles
Delay	122	41	13	1	70
Dose reduction	141	30	5	1	43

All toxicities and exposures could be combined in vectors. Denote by  $\mathbf{L}_k = (L_k^1, L_k^2, L_k^3, L_k^4, L_k^5)$  and  $\mathbf{A}_k = (A_k^1, A_k^2)$  the set of toxicities and exposures in cycle  $k$ , respectively. Over-bar is used to denote history, e.g.  $\bar{\mathbf{A}}_3 = (A_3^1, A_3^2, A_2^1, A_2^2)$ . Respectively, the symbols  $\mathcal{L}$  and  $\mathcal{A}$  represent the collection of all  $\mathcal{L}^j$  and  $\mathcal{A}^l$ , for  $j = 1, 2, \dots, 5$ , and  $l = 1, 2$ .

The logistic regression model for the effect of cumulative toxicities on HRe is as follows:

$$P(Y = 1 \mid \mathcal{L}, \mathcal{A}, s, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 \mathcal{L}^1 + \delta_2 \mathcal{L}^2 + \delta_3 \mathcal{L}^3 + \delta_4 \mathcal{L}^4 + \delta_5 \mathcal{L}^5 + \delta_6 \mathcal{A}^1 + \delta_7 \mathcal{A}^2 + \delta_8 s + \delta_9 \mathbb{1}_{\{\text{female}\}} + \delta_{10} \text{age}), \quad (4.19)$$

where  $s$  denotes the number of preoperative cycles.

The results from the estimation of Model (4.19) are presented in Table 4.8. Most of the cumulative toxicities have a negative effect on the probability of good HRe while subject-matter knowledge suggests that toxicities indicate therapy effectiveness and thus are expected to have a positive effect. However, all these effects are smaller than 0.2 on the log-odds scale which suggests that the sum of toxicity occurrences, treated as a continuous variable, is not the correct form to model the relationship between toxicities and the outcome. In addition, this model is biased because it corrects for time-dependent covariates, which are supposed to affect subsequent toxicities and as such should not be corrected for. However, this is an ad-hoc solution and is used only for exploratory purposes.

In order for toxicities to be confounders of the effect of exposure on the outcome, they should be associated with the exposure. These relationships could be assessed via pooled logistic regression models. In each of these models one component of the exposure is the dependent variable and is regressed on the toxicities from the preceding cycle and the baseline covariates. A pooled logistic regression is needed because it can easily incorporate the longitudinal nature of the processes.

The trial protocol prescribes allocation of cycle delays in case of haematological or oral toxicity. Therefore, we need to investigate the relationship between administration

Table 4.8: Toxicities effect estimates on histological response

	Estimate*	SE(Estimate)	p-value
(Intercept)	-1.648	0.732	0.024
Cumulative leucopenia	-0.150	0.276	0.587
Cumulative thrombocytopenia	0.248	0.516	0.631
Cumulative mucositis	-0.200	0.277	0.471
Cumulative infection	-0.113	0.230	0.625
Cumulative ototoxicity	-0.114	0.512	0.824
Cumulative cycle delay	0.478	0.318	0.132
Cumulative dose reduction	0.257	0.337	0.446
Number of preoperative cycles	0.470	0.356	0.187
Female	-0.392	0.354	0.267
Age	-0.028	0.029	0.338

\* log-odds

of cycle delays and WBC and platelets count, and mucositis. Such a model is defined as:

$$P(A_k^1 = 1 \mid L_{k-1}^1, L_{k-1}^2, L_{k-1}^3, A_{k-1}^1, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 L_{k-1}^1 + \delta_2 L_{k-1}^2 + \delta_3 L_{k-1}^3 + \delta_4 A_{k-1}^1 + \delta_5 \mathbb{1}_{\{\text{female}\}} + \delta_6 \text{age}). \quad (4.20)$$

As shown in Table 4.9, haematological toxicity increases the probability of administration of cycle delays, while the effect of thrombocytopenia is stronger than the one of leucopenia. A possible explanation for this huge difference in effect estimates is that thrombocytopenia occurs less often but when it is present the following cycle is most often delayed. The estimate of the effect of mucositis is negative and not significantly different from zero. An interesting discovery is the dependence of cycle delays in time. If the previous cycle was delayed, then it is more likely that also the current cycle will be delayed.

Table 4.9: Toxicities effect estimates on administration of delays

	Estimate*	SE(Estimate)	p-value
(Intercept)	-0.998	0.223	0.000
Leucopenia	0.151	0.458	0.743
Thrombocytopenia	1.197	1.021	0.241
Mucositis	-0.090	0.436	0.837
Delay of cycle $k-1$	1.604	0.497	0.001
Female	-0.087	0.307	0.777
Age	-0.015	0.024	0.524

\* log-odds

Before we can estimate the effects of toxicities on dose reductions, we should define the quantities that we are going to use. The study protocol prescribes dose reductions in case of simultaneous occurrence of infection and myelosuppression, or oral toxicity of grade 3 or 4. In both cases, with each occurrence of the symptoms the dose should be reduced by 20%. However, oral toxicity of grade 3 or 4 occurs in only 6 patient-cycles none of which is followed by a reduced dose cycle. This prevents us from estimating its effect on dose reduction. Further, CDDP should be discontinued after an episode of ototoxicity. Unfortunately, the protocol does not state how severe the latter should be. General subject-matter knowledge suggests that action is needed only in case of CTCAE grade 3 or 4. However, driven by the lack of such extreme quantities, we decide to use grades different from zero as indicators of ototoxicity. Discontinuations are also prescribed in case of neurological and cardiac toxicities but these are not developed during preoperative cycles and their effect cannot be estimated. As a result we can only evaluate the effects of myelosuppression and infection, and ototoxicity.

**Definition 4.13. Leucopenia for dose reduction**  $F_k^1$  is a dichotomous variable denoting the joint occurrence of infection and leucopenia throughout cycle  $k$ , i.e

$$F_k^1 = \mathbb{1}_{\{\text{infection and leucopenia in cycle } k\}} \begin{cases} 1 & , \text{ if infection CTCAE grade } \geq 1 \\ & \& \text{ WBC count CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & \text{ otherwise} \end{cases} \quad (4.21)$$

**Definition 4.14. Thrombocytopenia for dose reduction**  $F_k^2$  is a dichotomous variable denoting the joint occurrence of infection and thrombocytopenia throughout cycle  $k$ , i.e

$$F_k^2 = \mathbb{1}_{\{\text{infection and thrombocytopenia in cycle } k\}} \begin{cases} 1 & , \text{ if infection CTCAE grade } \geq 1 \\ & \& \text{ platelets count CTCAE grade } \geq 2 \text{ for cycle } k \\ 0 & \text{ otherwise} \end{cases} \quad (4.22)$$

**Definition 4.15. Ototoxicity**  $F_k^3$  is a discrete variable denoting the occurrence of ototoxicity throughout cycle  $k$ , i.e

$$F_k^3 = \mathbb{1}_{\{\text{ototoxicity in cycle } k\}} \begin{cases} 1 & , \text{ if ototoxicity CTCAE grade } \geq 1 \text{ for cycle } k \\ 0 & , \text{ if ototoxicity CTCAE grade } = 0 \text{ for cycle } k. \end{cases} \quad (4.23)$$

The model to assess the association of toxicities with dose reduction is as follows:

$$P(A_k^2 = 1 \mid \mathbf{F}_{k-1}, A_{k-1}^2, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 F_{k-1}^1 + \delta_2 F_{k-1}^2 + \delta_3 F_{k-1}^3 + \delta_4 A_{k-1}^2 + \delta_5 \mathbb{1}_{\{\text{female}\}} + \delta_6 \text{age}), \quad (4.24)$$

where  $\mathbf{F}_{k-1} = (F_{k-1}^1, F_{k-1}^2, F_{k-1}^3)$ .

Model (4.24) is a function of the toxicities from the previous cycle denoted by  $F_{k-1}^j$  for  $j = 1, 2, 3$  and  $k = 2, \dots, 6$ . Their effect estimates are corrected for the occurrence of dose reduction during the previous cycle (once applied reductions should be kept through subsequent cycles) and the baseline covariates.

The estimates are presented in Table 4.10. Leucopenia along with infection is associated with administration of dose reductions. On the contrary, thrombocytopenia and ototoxicity are estimated to have a small negative effect on dose reductions. As discussed before, these effect estimates are most probably driven by the imprecise definition of the side-effects. Low toxicity grades do not trigger therapy modifications but because of lack of severe toxicities we have to use a low cut-off CTCAE grade. The effect of previous cycle dose reduction is remarkable. This coincides with the recommendation in the protocol - once applied, dose reductions should be preserved at subsequent cycles.

Table 4.10: Toxicities effect estimates on administration of dose reductions

	Estimate*	SE(Estimate)	<i>p</i> -value
(Intercept)	-1.502	0.239	0.000
Leucopenia for dose reduction	0.591	0.682	0.386
Thrombocytopenia for dose reduction	-0.324	1.300	0.803
Ototoxicity	-0.260	1.005	0.796
Dose reduction of previous cycle	3.619	1.104	0.001
Female	-0.394	0.379	0.298
Age	0.018	0.027	0.508

\* log-odds

So far, we were assessing the confounding of the therapy modifications by toxicities. Equally important is the occurrence of feedback between the exposure and the potential confounders. We can investigate this kind of relationships by modelling toxicities throughout each cycle with exposure during the cycle. Since all decisions are made when a threshold of CTCAE grade is exceeded, we can model the occurrence of a toxicity that would justify an action. For that reason we build five pooled logistic regressions, one for each type of toxicity. Pooled logistic regression is used here because we assess the relation between exposure and toxicity within the same cycle conditional on the toxicities from the previous cycle. When a toxicity predicts both cycle delays and reductions, the effect of both exposures should be assessed. This applies to occurrences of leucopenia and thrombocytopenia. We estimate the effect of the exposures on each symptom of myelosuppression throughout a chemotherapy cycle by fitting model

$$P(L_k^j = 1 \mid \mathbf{A}_k, \mathbf{L}_{k-1}, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 A_k^1 + \delta_2 A_k^2 + \boldsymbol{\delta}_3^\top \mathbf{L}_{k-1} + \delta_4 \mathbb{1}_{\{\text{female}\}} + \delta_5 \text{age}), \quad (4.25)$$

where  $j = 1, 2$  and  $\mathbf{L}_{k-1}$  is the collection of all toxicities experienced in cycle  $k - 1$ , i.e.  $\mathbf{L}_{k-1} = (L_{k-1}^1, L_{k-1}^2, L_{k-1}^3, L_{k-1}^4, L_{k-1}^5)$ .

Oral toxicity in general affects both components of the exposure but as stated above, a severe mucositis that justifies administration of dose reductions is hardly observed. Hence, we predict episodes of oral toxicity based on cycle delays only by fitting model

$$P(L_k^3 = 1 \mid A_k^1, \mathbf{L}_{k-1}, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 A_k^1 + \boldsymbol{\delta}_2^\top \mathbf{L}_{k-1} + \delta_3 \mathbb{1}_{\{\text{female}\}} + \delta_4 \text{age}). \quad (4.26)$$

Infection and ototoxicity determine the administration of dose reductions. The effect of dose reductions on subsequent occurrence of infection or ototoxicity is modelled identically for the two, i.e. by estimating model

$$P(L_k^j = 1 \mid A_k^2, \mathbf{L}_{k-1}, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 A_k^2 + \boldsymbol{\delta}_2^\top \mathbf{L}_{k-1} + \delta_3 \mathbb{1}_{\{\text{female}\}} + \delta_4 \text{age}), \quad (4.27)$$

for  $j = 4, 5$ .

In the interest of presentation, Table 4.11 lists only the effects of the exposures on toxicities estimated by fitting Models (4.25), (4.26), and (4.27). Cycle delays and dose reductions are estimated to increase the probability of a subsequent haematological toxicity. The effects of cycle delays are very small. However, the correlation between dose reductions and haematological toxicity is much more pronounced. Their relationship is very complex mainly because of two reasons. First, dose reduction does not require myelosuppression during the preceding cycle. Such a side-effect could have occurred somewhere in the past but dose reduction is carried forward. Second, the study protocol warns that haematological toxicity will be cumulatively severe. Dose reductions are also more frequently administered in later cycle. Hence, this crude analysis, which corrects for time-dependent confounder, concludes that there is a strong positive association between dose reduction and subsequent myelosuppression. Fitting Models (4.26), and (4.27) concluded that cycle delays decrease the probability of oral toxicity, and dose reduction - those of infection and ototoxicity. Nonetheless, we are certain only of the direction of the association between infection and dose reduction.

Table 4.11: Effect estimates of exposure on subsequent occurrence of toxicities

Effect of	Effect on	Estimate*	SE(Estimate)	<i>p</i> -value
Cycle delay	Leucopenia	0.114	0.340	0.736
Dose reduction	Leucopenia	0.455	0.403	0.258
Cycle delay	Thrombocytopenia	0.070	0.600	0.907
Dose reduction	Thrombocytopenia	0.550	0.636	0.545
Cycle delay	Oral toxicity	-0.677	0.455	0.137
Dose reduction	Infection	-1.429	0.447	0.001
Dose reduction	Ototoxicity	-1.192	1.153	0.301

\* log-odds

The analyses in this section identified in the sample at hand some of the relationships advocated by clinical experts. However, all reported statistical tests do not have sufficient power to detect small effects. This is driven by the limited number of patients and the use of only 1 or 2 cycles per patient as these analyses were restricted to the preoperative chemotherapy data. Further, some of the presented models may be biased due to correcting for time-dependent confounders. Nevertheless, these models and the associated effect estimates are useful to recognise the complexity of the underlying processes. As such, their purpose is mainly exploratory. In the next section, we continue with building a MSM for the joint effect of cycle delays and dose reductions on HRe.

## 4.2.2 Model formulation and estimation

In this section we build a MSM for the joint causal effect of cycle delays and dose reductions. For this purpose we use cumulative exposures according to Definitions 4.11 and 4.12. However, the potential number of delayed cycles or cycles with reduced dose depends on the length of the preoperative treatment trajectory. Therefore, this aspect should be incorporated in the definition of the counter-factual outcomes.

**Definition 4.16. Counter-factual outcome** Let  $Y_{(\bar{a}^1, \bar{a}^2, s)}$  be the histological response arising from  $s$  preoperative cycles of chemotherapy, where  $(\bar{a}^1, \bar{a}^2, s) = ((a_2^1, a_2^2), (a_3^1, a_3^2), \dots, (a_s^1, a_s^2))$ , and  $a_k^j \in \{0, 1\}$ , for  $j = 1, 2$ , and  $k = 2, 3, \dots, s$ .

From Definition 4.16 it follows that there are four counter-factual outcomes in case of two preoperative cycles, which correspond to all combinations of cycle-two delay and dose reduction; 16 ( $2^4$ ) counter-factual outcomes after three preoperative cycles; 64 counter-factual outcomes after four preoperative cycles; 1024 counter-factual outcomes after six preoperative cycles. The number of outcomes grows exponentially with the number of preoperative cycles. We model in a parametric way how the probability of good counter-factual HRe depends on the number of delayed cycles, occurrences of dose reduction, and number of preoperative cycles as:

$$P(Y_{(\bar{a}^1, \bar{a}^2, s)} = 1) = \text{expit}(\beta_0 + \beta_1 \mathcal{A}^1 + \beta_2 \mathcal{A}^2 + \beta_3 s), \quad (4.28)$$

using 177 patient profiles, where  $\mathcal{A}^1$  and  $\mathcal{A}^2$  denote cumulative exposure according to Definitions 4.11 and 4.12, respectively.

Model (4.28) is a MSM for the counter-factual outcomes defined above. However, only one counter-factual outcome per patient is observed. This prevents us from estimating the parameters of Model (4.28) using the available data. Rather, we use a weighted regression

$$P(Y = 1 \mid \mathcal{A}, s) = \text{expit}(\gamma_0 + \gamma_1 \mathcal{A}^1 + \gamma_2 \mathcal{A}^2 + \gamma_3 s) \quad (4.29)$$

to model the factual outcomes of the patients.

When Model (4.29) is weighted by IPTWs, the  $\gamma$ -parameters are unbiased estimates of the  $\beta$ -parameters of Model (4.28). In order to estimate the IPTWs, we need to build models for allocation of therapy modifications. In principle, we have to model the joint

distribution of delays and reductions but this could be quite involving. In addition, there are few available software packages that can handle multivariate modelling. Alternatively, as shown in Section 2.2.1, we apply the conditional probability rule and model the two components of the exposure separately while conditioning the occurrence of one on the observation of the other. The natural temporal order of delays and reductions resolves possible issue on the decision on which to condition. Furthermore, we can improve the weights used in the model in Section 4.1.2 by estimating *stabilised weights*.

Therapy modifications are determined cycle-wise. Therefore, we also model the per cycle occurrence of an exposure of interest. For that purpose, we use a pooled logistic regression, which treats every patient-cycle as an independent observation. The occurrence of each exposure in cycle  $k$  is conditioned on the exposure in the previous cycle, the toxicities developed throughout the previous cycle, and the baseline confounders. In order to correctly incorporate the dependence between consecutive cycles, we introduce time into the models. The function of time and the overall intercept together determine the overall probability of exposure allocation.

A model for allocation of delays could be formulated as follows:

$$p_k^{del} = P(A_k^1 = 1 \mid \mathbf{A}_{k-1}, L_{k-1}^1, L_{k-1}^2, L_{k-1}^3, \mathbf{V}, k \leq s) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 A_{k-1}^1 + \alpha_3 A_{k-1}^2 + \alpha_4 L_{k-1}^1 + \alpha_5 L_{k-1}^2 + \alpha_6 L_{k-1}^3 + \alpha_7 \mathbb{1}_{\{\text{female}\}} + \delta_8 \text{age}), \quad (4.30)$$

for  $k = 2, \dots, 6$ , where  $\alpha_0$  and  $\alpha_1 k$  together form a cycle-dependent intercept, and both  $A_1^1$  and  $A_1^2$  are equal to zero by definition. We denote the probability of delaying cycle  $k$  by  $p_k^{del}$ . To construct stabilised weights we also need to build a model for allocation of delays but only based on exposure from the previous cycle. This probability we denote by  $p_k^{del,*}$ , and estimate from model

$$p_k^{del,*} = P(A_k^1 = 1 \mid \mathbf{A}_{k-1}, k \leq s) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 A_{k-1}^1 + \alpha_3 A_{k-1}^2). \quad (4.31)$$

Together the predicted probabilities from Models (4.30) and (4.31) allow us to construct the delays' component of the weights. Introducing an individual index  $i$ , subject-specific weights that break the dependence between toxicities and cycle delays, are computed as

$$SW_i^{del} = \prod_{k=2}^{s_i} \frac{[(p_{i,k}^{del,*})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{del,*})^{1-A_{i,k}^1}]}{[(p_{i,k}^{del})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{del})^{1-A_{i,k}^1}]}, \quad (4.32)$$

where  $s_i$  denotes the number of preoperative cycles of patient  $i$ .

The numerator of (4.32) is the probability of observed exposure given exposure history, while the denominator is the probability of observed exposure given treatment and covariate history.

In the same fashion, we build a model for allocation of dose reductions. The probability of dose reduction we condition on current cycle delay, i.e for  $k = 2, \dots, 6$

$$p_k^{red} = P(A_k^2 = 1 \mid A_k^1, \mathbf{A}_{k-1}, \mathbf{F}_{k-1}, \mathbf{V}, k \leq s) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 A_k^1 + \alpha_3 A_{k-1}^1 + \alpha_4 A_{k-1}^2 + \alpha_5 F_{k-1}^1 + \alpha_6 F_{k-1}^2 + \alpha_7 F_{k-1}^3 + \alpha_8 \mathbb{1}_{\{\text{female}\}} + \alpha_9 \text{age}). \quad (4.33)$$

Likewise, we define a model for reductions only based on exposure history, i.e.

$$p_k^{red,*} = P(A_k^2 = 1 \mid A_k^1, \mathbf{A}_{k-1}, k \leq s) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 A_k^1 + \alpha_3 A_{k-1}^1 + \alpha_4 A_{k-1}^2), \quad (4.34)$$

and compute the dose reduction component of the weights as

$$SW_i^{red} = \prod_{k=2}^{s_i} \frac{[(p_{i,k}^{red,*})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{red,*})^{1-A_{i,k}^1}]}{[(p_{i,k}^{red})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{red})^{1-A_{i,k}^1}]}. \quad (4.35)$$

However, the use of the product of the two components of the weights is not enough to account for all features of the data. What is left is to incorporate the effect of varying length of preoperative treatment trajectories. This problem could be approached in the same way as censoring in survival analysis. We observe preoperative treatment only until surgery takes place, i.e. surgery terminates preoperative treatment. Let us define a random variable  $C_k$  that takes value 0 for all cycles and value 1 when surgery is scheduled. As discussed in the introduction of Section 4.2, surgery is scheduled just like a subsequent course of treatment on the basis of the experienced toxicities. Therefore, if surgery was not scheduled, i.e. it will not be the next course of treatment,  $C_k$  is equal to 0. In fact,  $C_k$  is equal to 0 for all  $k < s$ . On the contrary, during the last preoperative cycle the planning of surgery takes place and  $C_s$  is equal to 1 ( $s$  denotes the number of preoperative cycles, as usual). Then a model for length of the preoperative treatment is as follows:

$$p_k^{ncyc} = P(C_k = 0 \mid \bar{\mathbf{C}}_{k-1} = \mathbf{0}, \mathbf{A}_k, \mathbf{L}_k, \mathbf{V}, k \leq s) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 A_k^1 + \alpha_3 A_k^2 + \alpha_4 L_k^1 + \alpha_5 L_k^2 + \alpha_6 L_k^3 + \alpha_7 L_k^4 + \alpha_8 L_k^5 + \alpha_9 \mathbb{1}_{\{\text{female}\}} + \alpha_{10} \text{age}), \quad (4.36)$$

where  $\mathbf{0}$  is a vector of zeros of length  $k - 1$ . Following the examples above, a model for the probability of surgery not following this cycle based on treatment history only is:

$$p_k^{ncyc,*} = P(C_k = 0 \mid \bar{\mathbf{C}}_{k-1} = \mathbf{0}, \mathbf{A}_k, k \leq s) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 A_k^1 + \alpha_3 A_k^2), \quad (4.37)$$

and the subject-specific weight is estimated according to

$$SW_i^{ncyc} = \frac{1 - p_{i,k=s_i}^{ncyc,*}}{1 - p_{i,k=s_i}^{ncyc}} \prod_{k=1}^{s_i-1} \frac{p_{i,k}^{ncyc,*}}{p_{i,k}^{ncyc}}. \quad (4.38)$$

In fact, by conditioning Equations (4.30) – (4.34) on  $k \leq s$  we implicitly condition on  $C$ , i.e. on  $\bar{\mathbf{C}}_{k-1} = \mathbf{0}$  as in Equations (4.36) and (4.37). The final subject-specific weight is calculated as

$$SW_i = SW_i^{del} \times SW_i^{red} \times SW_i^{ncyc}, \quad (4.39)$$

a product of the distinct components of the weights.

For completeness of exposition, we should note that the components of the final weight, i.e.  $SW_i^{del}$ ,  $SW_i^{red}$ , and  $SW_i^{ncyc}$ , cannot all be estimated for everyone. In the current example the only two exceptions are the two patients with a single preoperative chemotherapy course. As explicitly specified, Equations (4.30) and (4.33) are only defined for  $k \geq 2$ . They need toxicity from previous cycle, and there are no toxicities from cycle before the first cycle. In addition, during the first cycle always full dose is given and the first cycle cannot be delayed by definition. Therefore, we model therapy modifications from cycle 2 onwards. As a result  $SW_i^{del}$  and  $SW_i^{red}$  cannot be estimated for patients who underwent only one preoperative cycle. In fact,  $SW_i^{del}$  and  $SW_i^{red}$  are both equal to one when  $s_i = 1$ . In other words, the final weight of the two patients with  $s_i = 1$  is equal to  $SW_i^{ncyc}$ .

The probabilities that comprise the weights are estimated cycle-wise, while finally subject-specific weights are used, i.e. one weight per patient. Table 4.12 presents the distribution of the probabilities of observed exposure, subject-specific components of the weights, and the final weights. Approximately half of the delays (or lack of delay) are correctly predicted (median probability of observed delay allocation is 0.729, row 1 of Table 4.12). On a subject rather than cycle level this results in cycle delay component with mean equal to 1. Less dose reductions than cycle delays are observed during the preoperative courses. That is why at least 75% of the probabilities of observed dose reduction administration are high. For this reason, the range of  $\hat{SW}_i^{red}$  is smaller than that of  $\hat{SW}_i^{del}$ . The probabilities of observed surgery timing are most extreme and left skewed which results in higher variation of the corresponding weights. Some fitted probabilities are exactly 0 or 1. In principle, this could indicate possible violation of the positivity assumption, however, we do not dive into this now because the current MSM is used for illustrative purposes. The final subject-specific weights have a mean equal to 0.996 which is a desirable property. However, the distribution of the weights should further be examined. Figure 4.4 shows that the estimated weights are symmetrically distributed around their mean which is best observed on the logarithmic scale. This applies especially for the final weights which we are most interested in.

### 4.2.3 Results and discussion

The causal and associational effects of cycle delay, dose reductions, and length of preoperative treatment on HRe are presented in terms of odds-ratios in Table 4.13. An additional administration of cycle delay is estimated to cause 1.954 times more often a GR than a poor one. This result contradicts researchers' hypothesis that delays should allow the body and tumour to recover and possibly the tumour to regrow. These results

Table 4.12: Distribution of estimated probabilities of observed exposure and weights

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\hat{P}(A_{(i, k)}^1 = a_{(i, k)}^1), \text{ for } k \geq 2$	0.177	0.381	0.729	0.608	0.753	0.876
$\hat{S}\hat{W}_i^{del}$	0.432	0.971	0.994	0.999	1.020	1.651
$\hat{P}(A_{(i, k)}^2 = a_{(i, k)}^2), \text{ for } k \geq 2$	0.118	0.773	0.824	0.735	0.868	1.000
$\hat{S}\hat{W}_i^{red}$	0.485	0.957	0.998	1.000	1.025	1.468
$\hat{P}(S_{(i, k)} = s_{(i, k)}), \text{ for } k \geq 1$	0.000	0.615	0.881	0.754	0.931	1.000
$\hat{S}\hat{W}_i^{ncyc}$	0.255	0.884	0.975	0.996	1.080	2.671
$\hat{S}\hat{W}_i$	0.130	0.860	0.965	0.996	1.071	2.926

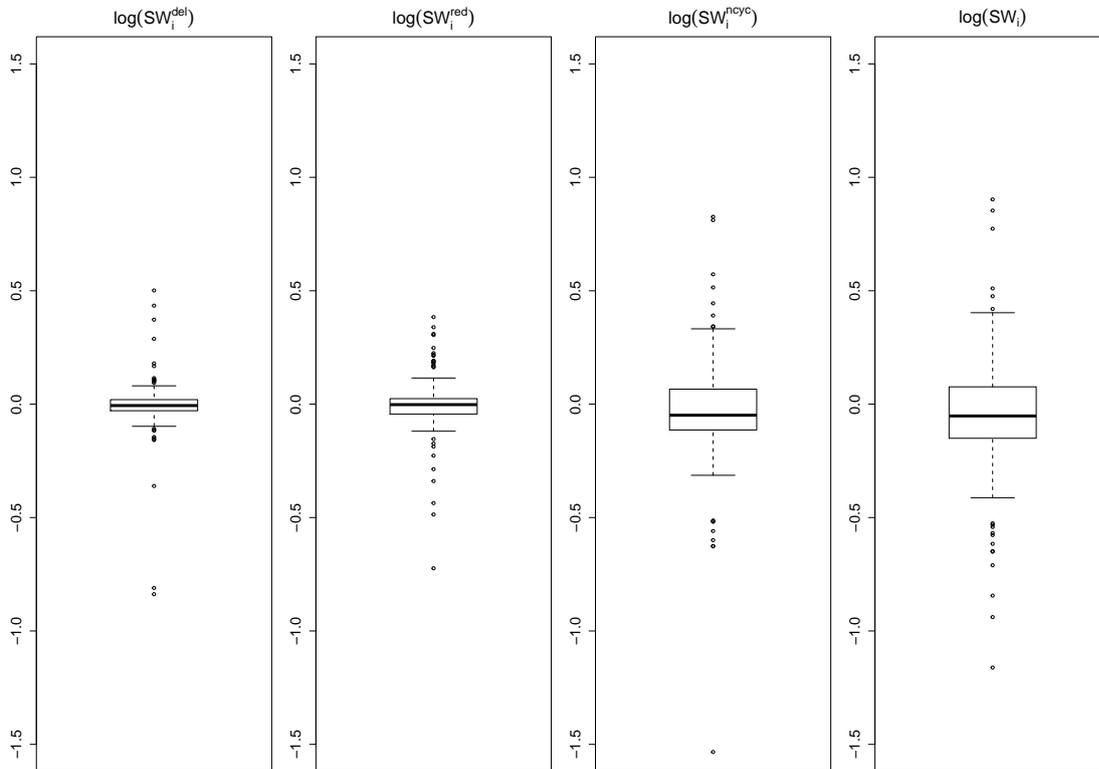


Figure 4.4: Boxplots of subject-specific weights and its components

suggest that cycle delays due to toxicity might have a positive effect on HRe. This is based on the notion that toxicities indicate that the body, and thus the tumour are

suffering. In fact, our definition of cycle delay captures a great amount of small delays. Some of these probably just represent the between-patient variation of the time that the body needs to recover from the cytotoxic drugs, and as such are essential for patients' well-being. Administration of dose reductions and an additional preoperative cycle are also positively associated with GR. Yet, there are very few occurrences of dose reduction during the preoperative cycles that does not allow us to gain confidence in the direction of its relationship with HRe. An extra preoperative cycle is estimated to cause on average a 45.5% increase in the frequency of GRs. This result is in line with researchers' expectations.

Association effect estimates of therapy modifications on HRe are less extreme than the causal ones. Yet, they indicate the same type of relations between therapy modifications and HRe. When the indirect effect of delays and reductions on HRe through toxicities is eliminated, the interdependences between exposure and outcome are enhanced. The presence of difference, although very mild, between causal and association parameter estimates indicate that indeed the effects of therapy modifications are confounded, at least to some extent, with toxicities. The validity of the estimated weights, which the causal effect estimates are highly dependent on, is confirmed by the mean of the weights being equal to 0.996 (theory expects the mean to be equal to one).

Table 4.13: Causal and association odds-ratio estimates for the effects of cycle delay, dose reductions, and length of preoperative treatment on histological response

	OR	95% CI	<i>p</i> -value
Causal: cycle delay	1.954	[1.045 : 3.651]	0.037
Causal: dose reduction	1.432	[0.746 : 2.749]	0.282
Causal: number of preoperative cycles	1.455	[0.749 : 2.826]	0.270
Association: cycle delay	1.613	[0.882 : 2.952]	0.123
Association: dose reduction	1.287	[0.704 : 2.353]	0.413
Association: number of preoperative cycles	1.463	[0.793 : 2.698]	0.225

In this section we discussed how difficult the application of MSMs with chemotherapy data can be. However, the model built in this section is still not able to capture all features of the data. That is why the next section presents the final MSM to estimate the causal effects of therapy modifications on HRe. The next model upgrades the one presented here. First, it includes one more exposure - surgery delay in terms of days since anticipated surgery day. Its effect is of interest for the researchers because surgery delays, in contrast to cycle delays, are hardly driven by toxicities. Rather, most of them are purely administrative delays. Second, one more type of weights will be introduced to account for possible selection bias due to analysis of patient records with available HRe only. And last but not least, next section devotes special attention to selection of the models for exposure allocation that comprise the weights. After all, the validity of the inference through MSMs rests upon lack of model misspecification of any of the models used in the main analysis, i.e. the models for exposure allocation and the MSM itself. This was withheld until now because the models presented so far in this chapter were

build for illustrative purposes while model selection is a timely and complex process.

### 4.3 MSM for joint causal effect of cycle and surgery delays, and dose reductions on histological response

Section 4.1 shows how to build and estimate a MSM for the causal effect of chemotherapy cycle delays on HRe. Section 4.2 extends this model to additionally assess the effect of dose reductions on HRe. In addition, it enables the use of all patient records by allowing variable preoperative treatment length. Yet, this model could further be improved. First, we want to investigate the effect of surgery delay on HRe. Second, we believe it is necessary to adjust for possible selection bias due to analysis of patient records with available HRe. Further, the models presented in the previous sections do not address the models for exposure allocation from the perspective of the assumptions of the methodology. The validity of the inference drawn upon a MSM relies on correct model specification of both – the models for exposure allocation and the MSM itself. However, they used the most straightforward model formulations, e.g. linear functions of time and age, and no interaction terms. This section devotes special attention to the model selection challenge.

Our understanding of the interdependencies between exposures, time-dependent confounders, and HRe can easily be extended to incorporate surgery delay. Delay of surgery could be regarded as delay of subsequent treatment just like cycle delays. Although 34.7% of the surgeries were delayed due to administrative reasons which cannot be predicted, 15.8% were delayed because of haematological toxicity and 2.5% due to other toxicities (see Figure 3.9b). In addition, the trial protocol suggest that surgery could be delayed because of the same reasons that motivate cycle delays. The DAG on Figure 4.2 needs only to be extended with an additional node of surgery delay just before HRe vertice to adopt the additional exposure (see Figure 4.5). Surgery delay is expected to affect HRe by decreasing the probability of a good response thus an edge extends from *surgery delay* to *histological response*. Since some of the reasons for surgery delay are toxicities, a directed edge connects the toxicities from the last preoperative cycle to *surgery delay*. Similarly, the exposure from the cycle when surgery was scheduled could also directly affect surgery delay. This is represented by an edge between *exposure* and *surgery delay*. And last but not least, before a the exposure and confounders could influence surgery delay, surgery should be scheduled, i.e. *scheduled surgery* nodes transmit these effects to *surgery delay*.

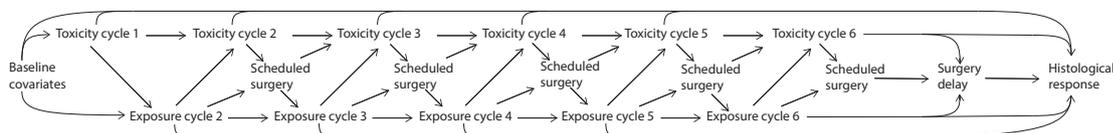


Figure 4.5: DAG describing the causal relationships between exposures, confounders and histological response

### 4.3.1 Investigation of surgery delay confounding

Before incorporating surgery delay into the MSM in the same fashion as the other exposures, we should confirm its confounding with the last preoperative cycle toxicities. This effect could be assessed through a logistic regression of surgery delay indicator on toxicities. Let us first define surgery delay.

**Definition 4.17. Surgery delay**  $A^3$  is a dichotomous variable denoting surgery delay, i.e.

$$A^3 = \mathbb{1}_{\{\text{surgery delay} \geq 7 \text{ days}\}} \begin{cases} 1 & , \text{ if it is performed later than day 27 of the last preoperative cycle} \\ 0 & , \text{ if it is performed before day 28 of the last preoperative cycle.} \end{cases} \quad (4.40)$$

According to this definition the surgery of 76 patients has been delayed.

A model for occurrence of surgery delay given leucopenia, thrombocytopenia, mucositis, and the baseline confounders is:

$$P(A^3 = 1 \mid L_s^1, L_s^2, L_s^3, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 L_s^1 + \delta_2 L_s^2 + \delta_3 L_s^3 + \delta_4 \mathbb{1}_{\{\text{female}\}} + \delta_5 \text{age}). \quad (4.41)$$

Table 4.14: Effect of last preoperative cycle toxicities on surgery delay

	Estimate*	SE(Estimate)	<i>p</i> -value
(Intercept)	-0.186	0.462	0.688
Leucopenia	-0.383	0.397	0.334
Thrombocytopenia	0.052	0.656	0.937
Mucositis	-1.059	0.474	0.025
Female	-0.105	0.320	0.744
Age	0.012	0.024	0.614

\* log-odds

The estimates of  $\delta$ -parameters of Model (4.41) are presented in Table 4.14. Thrombocytopenia is not associated with surgery delays. However, occurrence of mucositis and leucopenia are associated with shorter surgery delays or lack of surgery delay. These results are not as expected. However, they are most probably explained by the fact that hardly 18% of the surgeries were delayed due to toxicities. Among the rest, the patients did not have any of the three toxicities, and as a result we obtain negative association parameters. Yet, the results prove that the effect of surgery delay is confounded with the effect of oral toxicity on HRe. Therefore an inverse-probability-of-treatment weights should be estimated to correct for this confounding.

### 4.3.2 Model formulation and estimation

The inclusion of surgery delay indicator in the exposure doubles the number of counter-factual outcomes. We need to redefine counter-factual outcome.

**Definition 4.18. Counter-factual outcome** Let  $Y_{(\bar{a}^1, \bar{a}^2, a^3, s)}$  be the histological response arising from  $s$  preoperative cycles of chemotherapy, where  $(\bar{a}^1, \bar{a}^2, a^3, s) = ((a_2^1, a_2^2), (a_3^1, a_3^2), \dots, (a_s^1, a_s^2), a^3)$ , and  $a_k^j \in \{0, 1\}$ , for  $j = 1, 2, 3$ , and  $k = 2, 3, \dots, s$ .

A MSM for the joint causal effect of cycle delays, dose reductions, and surgery delay on HRe is as follows:

$$P(Y_{(\bar{a}^1, \bar{a}^2, a^3, s)} = 1) = \text{expit}(\beta_0 + \beta_1 \mathcal{A}^1 + \beta_2 \mathcal{A}^2 + \beta_3 A^3 + \beta_4 s), \quad (4.42)$$

where  $\mathcal{A}^1$  and  $\mathcal{A}^2$  denote cumulative exposure (for the whole preoperative treatment trajectory) according to Definitions 4.11 and 4.12, respectively.

Since only one of the counter-factual outcomes is observed for each patient we cannot directly estimate Model (4.42). Instead, we use a weighted regression model

$$P(Y = 1 \mid \mathcal{A}, A^3, s) = \text{expit}(\gamma_0 + \gamma_1 \mathcal{A}^1 + \gamma_2 \mathcal{A}^2 + \gamma_3 A^3 + \gamma_4 s), \quad (4.43)$$

the factual analogue of Model (4.42).

In order to fit Model (4.42) we need to estimate inverse-probability-of-treatment weights. This is done separately for each component of the exposure. As stated at the beginning of this section, the fallacy of the previous MSMs is that they do not address the issue with model building for allocation of exposure. This is done here by starting with the model for administration of cycle delays. An important aspect of the use of pooled logistic regression for longitudinal analysis is the correct incorporation of time into the model. This is usually done by modelling time-dependent intercept. Table 4.15 presents six alternative formulations of this model. Specification 1 is the one that was used in the previous MSMs. It comprises of linear functions of cycle number and patient's age at registration; indicators of previous cycle delay and dose reduction; indicators of leucopenia, thrombocytopenia, and mucositis of grade two or higher; and gender indicator. The mean of the weights estimated using this model is 0.999. The effect of cycle delays on HRe is estimated through fitting Model (4.43) on a pseudopopulation created by the weights correcting for confounding of cycle delays with toxicities only. However, linear function of cycle number is very restrictive. That is why in specification 2 a quadratic term of cycle number is added. This extension is justified by 7 units decrease in the Akaike's information criterion (AIC). Nevertheless, there is no biological reasoning behind the quadratic function of time, and this made us think that the model goodness-of-fit increases just because of the freedom that the additional term allows to the model. As a further improvement we created indicator variables for each cycle number, and thus removed all possible dependence between the time-dependent intercepts of the model. However, there are very few observations with more than two preoperative cycles, and this disables us to accurately estimate cycle indicator parameters. Unexpectedly this resulted in poorer goodness-of-fit (AIC increases).

Table 4.15: Models for allocation of cycle delays

	Specification	AIC†	Mean weight (SD†)	Min/Max	Estimate* (SE†)
1	Linear functions of cycle and age, indicators of previous cycle delay and dose reduction, leucopenia, thrombocytopenia, mucositis, and gender	281.3	0.999 (0.112)	0.433/1.651	0.641 (0.331)
2	As per model 1 plus quadratic term of cycle number	274.4	0.999 (0.105)	0.431/1.661	0.629 (0.333)
3	As per model 1 but cycle number indicators instead of linear function	278.1	0.999 (0.103)	0.436/1.603	0.632 (0.333)
4	As per model 2 but 3-knot RCS† for age instead of linear function	275.9	1.000 (0.121)	0.427/1.693	0.609 (0.335)
5	As per model 4 but including interaction between age and gender	272.9	0.994 (0.202)	0.347/1.901	0.839 (0.343)
6	As per model 2 but including interaction between age and gender	273.1	0.999 (0.166)	0.403/1.650	0.750 (0.341)

\* log-odds

† AIC, Akaike's information criterion; SD, standard deviation; SE, standard error; RCS, restricted cubic spline.

Table 4.16: Models for allocation of dose reductions

	Specification	AIC†	Mean weight (SD†)	Min/Max	Estimate* (SE†)
1	Linear functions of cycle and age, indicators of current cycle delay, previous cycle delay and dose reduction, joint occurrence of leucopenia and infection, thrombocytopenia, and infection, mucositis, and gender	212.6	1.000 (0.114)	0.485/1.468	0.040 (0.335)
2	As per model 1 but separate indicators of leucopenia, thrombocytopenia and infection	199.8	1.004 (0.369)	0.404/3.089	0.129 (0.430)
3	As per model 2 but including interaction terms between leucopenia and infection, and thrombocytopenia and infection	203.7	1.004 (0.370)	0.406/3.110	0.121 (0.428)

\* log-odds

† AIC, Akaike's information criterion; SD, standard deviation; SE, standard error.

Initially, a linear function of age was used. In order to allow more freedom to the model, the linear function of age was substituted with a three-knot Restricted cubic spline (RCS). RCS is a spline function that is restricted to linearity at both ends. This modification is added to specification 2. However, it did not result in a better fit. There is a biological reasoning behind the hypothesis that the effect of age might depend on gender. Cell growth and respectively recovery varies with age. That is why specifications 5 and 6 include interaction terms between RCS and linear function of age and gender, respectively. The fifth model resulted in very different effect on HRe estimate and no further use of the model was done. Quadratic function of age was also tested but did not lead to a better fit. Finally, because of the biological justification of the interaction between age and gender, we decided to use specification 6. It is mathematically expressed as:

$$P(A_k^1 = 1 | A_{k-1}, L_{k-1}^1, L_{k-1}^2, L_{k-1}^3, \mathbf{V}) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 k^2 + \alpha_3 A_{k-1}^1 + \alpha_4 A_{k-1}^2 + \alpha_5 L_{k-1}^1 + \alpha_6 L_{k-1}^2 + \alpha_7 L_{k-1}^3 + \alpha_8 \mathbb{1}_{\{\text{female}\}} + \alpha_9 \text{age} + \alpha_{10} \text{age} * \mathbb{1}_{\{\text{female}\}}), \quad (4.44)$$

where  $k$  is the cycle number taking values from 2 to 6, and  $\text{age}$  is a continuous variable centred at 16 years.

Next we discuss the model for allocation of dose reductions. The same improvements upon the functional forms of cycle number, age, and gender were tested but none increased the model goodness-of-fit. The previous formulation of the model for administration of dose reductions uses indicators of joint occurrence of leucopenia and infection, and thrombocytopenia and infection. The possibility to split those into separate indicators of types of myelosuppression and infection was considered (see specification 2 in Table 4.16). This alternative resulted in much better fit and the effect estimate of dose reductions on HRe moved away from zero. However, the protocol prescribes dose reductions when both haematological toxicity and infection occur. For this reason we tried to augment specification 2 with interaction terms between the two indicators of myelosuppression and infection. The distribution of the resulted weights did not really change while the model goodness-of-fit became worse. As a result, we decide to use specification 2. It is expressed as:

$$P(A_k^2 = 1 | A_k^1, \mathbf{A}_{k-1}, L_{k-1}^1, L_{k-1}^2, L_{k-1}^{4,*}, L_{k-1}^5, \mathbf{V}) = \text{expit}(\alpha_0 + \alpha_1 k + \alpha_2 A_k^1 + \alpha_3 A_{k-1}^1 + \alpha_4 A_{k-1}^2 + \alpha_5 L_{k-1}^1 + \alpha_6 L_{k-1}^2 + \alpha_7 L_{k-1}^{4,*} + \alpha_8 L_{k-1}^5 + \alpha_9 \mathbb{1}_{\{\text{female}\}} + \alpha_{10} \text{age}), \quad (4.45)$$

where  $L_k^1, L_k^2$ , and  $L_k^5$  indicate occurrences of leucopenia, thrombocytopenia, and ototoxicity according to Definitions 4.6, 4.7, and 4.10, respectively, and  $L_k^{4,*}$  indicates occurrence of infection of CTC AE grade larger than 0.

One might question why toxicities from two cycles afore are not included in these models, or whether indicator variables are enough to capture the effect of toxicities on exposure. First, exposure is modified cycle-wise based on the most current toxicities.

Second, the presence of exposure from the previous cycle could explain the current cycle delay in absence of toxicities as the protocol states that once reduced the dose should be kept constant until subsequent occurrence of extreme toxicities. More extreme toxicities might lead to a larger dose reduction but our exposure does not distinguish between 20 and 40% reduction (according to the protocol, dose reductions are applied in steps of 20%), and thus will not be affected by more extreme values.

Similar to the models for allocation of cycle delays and dose reductions, different functional forms of the model for administration of surgery delays were tested. However, only interaction between gender and linear function of age improved the model fit. The final model used to compute the surgery delay component of the weights is:

$$P(A^3 = 1 \mid \mathbf{A}_s, L_s^1, L_s^2, L_s^3, s, \mathbf{V}) = \text{expit}(\alpha_0 + \alpha_1 s + \alpha_2 A_s^1 + \alpha_3 A_s^2 + \alpha_4 L_s^1 + \alpha_5 L_s^2 + \alpha_6 L_s^3 + \alpha_7 \mathbb{1}_{\{\text{female}\}} + \alpha_8 \text{age} + \alpha_9 \text{age} * \mathbb{1}_{\{\text{female}\}}), \quad (4.46)$$

where  $s$  stands for the number of preoperative cycles, and  $L_k^1, L_k^2$ , and  $L_k^3$  indicate occurrences of leucopenia, thrombocytopenia, and oral toxicity according to Definitions 4.6 – 4.8, respectively.

The last component of the weights takes into account the differences in preoperative treatment length. In the previous section we defined a scheduled surgery indicator variable  $C_k$  that marks after which cycle the surgery is planned to take place. Among the tested alternatives, the model that has the best fit characterises with cycle-specific intercept, the two sets of toxicities (for cycle delays and dose reductions defined in Section 4.2), linear function of age, and gender indicator. To be more precise about the time-dependent intercepts, a common intercept is estimated for the patients with surgery after cycle 4 or later. This estimate is based on 10 patient records. The model is as follows:

$$P(C_k = 1 \mid \bar{\mathbf{C}}_{k-1} = \mathbf{0}, \mathbf{A}_k, \mathbf{L}_k, \mathbf{V}) = \text{expit}(\alpha_0 + \alpha_1 \mathbb{1}_{\{\text{second cycle}\}} + \alpha_2 \mathbb{1}_{\{\text{third cycle}\}} + \alpha_3 \mathbb{1}_{\{\text{cycle four, five or six}\}} + \alpha_4 A_k^1 + \alpha_5 A_k^2 + \alpha_6 L_k^1 + \alpha_7 L_k^2 + \alpha_8 L_k^3 + \alpha_9 F_k^1 + \alpha_{10} F_k^2 + \alpha_{10} F_k^3 + \alpha_{11} \mathbb{1}_{\{\text{female}\}} + \alpha_{12} \text{age}), \quad (4.47)$$

where  $L_k^1, L_k^2$ , and  $L_k^3$  indicate occurrences of leucopenia, thrombocytopenia, and oral toxicity according to Definitions 4.6 – 4.8, respectively, and  $F_k^1, F_k^2$ , and  $F_k^3$  indicate confounders of dose reduction according to Definitions 4.13 to 4.15.

Models (4.44) – (4.47) estimate cycle-wise the probability of exposure allocation and surgery schedule conditional on exposure and covariate history. These probabilities could be used to estimate the probability of observed exposure. This is done by using binomial likelihood approach,  $[P(Z = 1)^z] \times [(1 - P(Z = 1))^{(1-z)}]$ , where  $Z$  is a dichotomous random variable, and  $z$  denotes its realisation, i.e.  $z \in \{0, 1\}$ . Then, the probability of the whole preoperative exposure history is estimated as a product of the cycle-dependent probabilities of observed exposure. However, for the estimation of the stabilised inverse-probability-of-treatment weights we also need the probability of observed exposure given

treatment history alone. These are estimated separately for each component of the weights by using only the cycle and exposure components of Models (4.44) – (4.47), respectively, as shown in Section 4.2.2.

The only new component of the exposure, and respectively the weights, is surgery delay. The probability of surgery delay given exposure and covariate history is estimated through Equation (4.46). Its complement estimating the probability of surgery delay only based on exposure history is estimated through fitting model  $P(A^3 = 1 | s, \mathbf{A}_s)$ . Then the probability of observed surgery timing is equal to

$$SW_i^{srg} = \frac{P(A^3 = 1 | s, \mathbf{A}_s)}{P(A^3 = 1 | s, \mathbf{A}_s, L_s^1, L_s^2, L_s^3, \mathbf{V})} \quad (4.48)$$

for patients with delayed surgery and to

$$SW_i^{srg} = \frac{1 - P(A^3 = 1 | s, \mathbf{A}_s)}{1 - P(A^3 = 1 | s, \mathbf{A}_s, L_s^1, L_s^2, L_s^3, \mathbf{V})} \quad (4.49)$$

for patients with surgery on time.

The final subject-specific weights are computed as a product of the weights for each component of the exposure, i.e.

$$SW_i = SW_i^{del} \times SW_i^{red} \times SW_i^{srg} \times SW_i^{ncyc}. \quad (4.50)$$

The distribution of each component and the final weights is visualised through box-plots on Figure 4.6. Logarithmic transformation is used to stretch the distribution of values smaller than one. Most of the elements of the weights are symmetrically distributed around their mean, approximately equal to zero. The final weights have larger range than any of its components, and more than half of the weights are smaller than 1 (the median indicated by the bold horizontal line within the box is below 0).

One more aspect of the data that we have not addressed so far is the missing values for HRe; for 17% of the patients HRe is not known. Some of them also lack surgery information which suggests that surgery was not performed at all. Those patients have the same age and gender distribution as bad responders. This is an indication that the patients we exclude from the analyses so far probably would have had a PR if operated at all. Therefore, an analysis of the complete cases probably introduces selection bias since only part of the bad responders are analysed. In order to correct for this possible selection bias we can estimate IPCWs where observations without HRe are censored prior to obtaining the results from the resected specimen. The general approach to such a selection bias is to build a model for the probability of observing HRe based on all patients (with missing and available outcome), and weight patients with observed HRe in the main analysis with the corresponding probability-of-observed-HRe weight.

Total treatment trajectories of these patients vary between only one cycle (4 patients) to six cycles (23 patients), where 38 patients lack HRe. However, date of surgery is often missing for those patients and thus we cannot identify the preoperative cycles and calculate their exposure values for this period. That is why we can estimate the IPCW only based on baseline patient profile. Tumour characteristics like location in

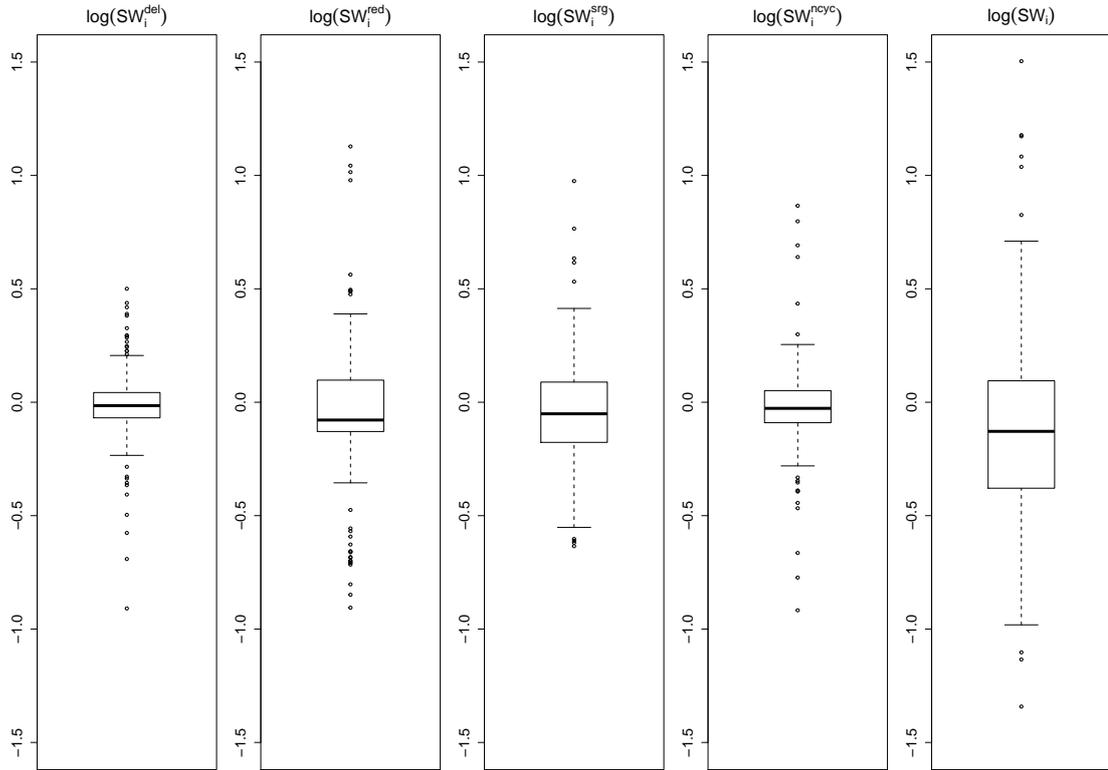


Figure 4.6: Boxplots of the final weights and all its components

the body and on the bone are not associated with missing HRe, therefore we build a censoring model using patients' age and gender. Further, the inability to calculate their preoperative exposure disables us to compute stabilised weights. We found that information on HRe highly depends on the country where the patient was treated. For example, the majority of patients come from hospitals in the UK and hardly anyone lacks HRe record. On the contrary, the three patients from Portugal all have missing HRe. Table 4.17 lists the number of patients with recorded and missing HRe by country of treatment for 219 patients. These are all patients randomised to the control arm but those without treatment data and who used G-CSF. For concreteness, these are the patients marked as 'complete cases based on independent variables (202)' on Figure 3.16 plus the 17 patients with unknown date of surgery. The latter were included for analysis of missing HRe because among them there are patients with available HRe as well as with missing HRe. If we include country indicator variables into a logistic regression model for missing HRe, the effect estimates for Canada, France, Portugal, Slovenia, and South Africa will be very extreme and unstable because of the small sample sizes and the one-sided observations (all missing vs. all available). As a result, some of the predicted probabilities of observed HRe will be equal to exactly 0 or 1. In order to stabilise the weights, we group the countries with a vary limited number of patients (Canada, France,

Portugal, Slovenia, and South Africa) into one category that we refer to as ‘Other’, and use this recoded country variable to estimate the probabilities of observed HRe.

Table 4.17: Distribution of recorded and missing HRe by country of treatment

HRe	Argentina	Belgium	Canada	Chile	Denmark	France
available	16	6	1	4	7	0
missing	3	2	0	5	2	1
	Portugal	Saudi Arabia	Slovenia	South Africa	The Netherlands	UK
available	0	17	0	0	29	101
missing	3	9	1	2	7	3

Let us define a random variable  $H$  taking value 1 if HRe was recorded, and 0 if not. Then a model for availability of HRe is defined as follows:

$$\begin{aligned}
 P(H = 1 \mid \mathbf{V}, \text{country}) = \text{expit}(\alpha_0 + \alpha_1 \mathbb{1}_{\{\text{female}\}} + \alpha_2 \text{age} + \alpha_3 \mathbb{1}_{\{\text{Belgium}\}} + \alpha_4 \mathbb{1}_{\{\text{Chile}\}} \\
 + \alpha_5 \mathbb{1}_{\{\text{Denmark}\}} + \alpha_6 \mathbb{1}_{\{\text{Other}\}} + \alpha_7 \mathbb{1}_{\{\text{Saudi Arabia}\}} + \alpha_8 \mathbb{1}_{\{\text{The Netherlands}\}} + \alpha_9 \mathbb{1}_{\{\text{UK}\}}),
 \end{aligned}
 \tag{4.51}$$

where  $\text{age}$  is a continuous variable centred at 16 years, and the reference country is Argentina.

Alternative formulations of Model (4.51) were also considered. For example various non-linear functions of age and possible interactions with gender. However, none of them increased the goodness-of-fit of the model. The estimated probabilities for observing patient’s HRe range between 0.108 and 0.979, which indicates good discriminative ability. The IPCW are then computed as the inverse of the fitted by Model (4.51) probabilities. The IPCW of the patients with known date of surgery and HRe (the 177 patients that we use to model their HRe) range between 1.021 and 7.487. We noted that the maximum IPCW corresponds to the patient from Canada, who has available HRe but takes part in the ‘Other’ country category, which compiles all countries with no available HRe. When we exclude this patient, the maximum IPCW becomes 2.197, in the next section we discuss the sensitivity of the results to the weight for the patient from Canada.

### 4.3.3 Results and discussion

The model presented in this section attempts to take all problems and data specifics into account. It assesses the joint causal effect of dose reductions, cycle and surgery delays during preoperative chemotherapy on HRe. Because the exposures depend on the length of the preoperative therapy, additionally the effect of an extra preoperative cycle is quantified. Table 4.18 lists the causal and association odds-ratio estimates of the four-variate exposure on HRe. The causal effect estimates were obtained using weights that are a product of  $SW_i$  according to Equation (4.50) and the inverse-probability-of-available-HRe estimated by fitting Model (4.51). There is a substantial difference between the

parameter estimates of the MSM and the unweighed analysis for all elements of the exposure but surgery delay. This might suggest that surgery delay is not confounded with toxicities although toxicity was indicated as the major reason for surgery delay in 18.3% of the delayed surgeries. Surgery delay, in contrast to cycle delay, has a negative causal effect on GR, i.e. a good HRe is 2.39 times less likely when surgery is delayed compared to surgery on time. There are two explanations for this discrepancy. First, most surgeries are delayed due to administrative reasons. This means that the patient was ready to undergo surgery on time but there was no surgeon and/or surgical theatre available. While this is the case, the patient further recovers from the side effects of the cytotoxic drugs and the tumour regrows. Such a delay decreases the probability of a GR. On the contrary, most cycles are delayed due to toxicities. During such a delay the patient and, respectively, the tumour continue to suffer. As a result, cycle delays are associated with GR. Second possible explanation lies in the definitions of the two delays. Cycle delays address short delays of three or more delays, while surgery is considered delayed only when it takes place one week or more after the anticipated date. As such short delays, i.e. cycle delays, allow the body to recover, while longer delays, i.e. surgery delay, leave a room for tumour regrowth besides body recovery. Thus the latter are associated with a PR. These results are further elaborated on in Chapter 6.

Table 4.18: Causal and association odds-ratio estimates for the joint effect of cycle delays, dose reductions, surgery delay, and the number of preoperative cycles on histological response

	OR	95% CI	<i>p</i> -value
Causal: cycle delay	2.567	[1.197 : 5.506]	0.016
Causal: dose reduction	1.752	[0.658 : 4.664]	0.263
Causal: surgery delay	0.419	[0.188 : 0.937]	0.036
Causal: number of cycles	1.056	[0.496 : 2.246]	0.888
Association: cycle delay	1.862	[0.968 : 3.581]	0.064
Association: dose reduction	1.225	[0.654 : 2.296]	0.527
Association: surgery delay	0.371	[0.183 : 0.754]	0.007
Association: number of cycles	1.440	[0.782 : 2.649]	0.243

The more exposures are incorporated into the MSM, the more extreme the causal effect of an additional cycle delay becomes. This indicates that the effects of different therapy modification approaches should really be corrected for one-another. The same applies for the effect of dose reductions on HRe. Unexpectedly, an additional preoperative chemotherapy course has a very small positive causal effect on the frequency of good HRe. An additional preoperative chemotherapy course is associated with 44% more frequent GR but is estimated to cause only 5.6% increase in the frequency of good response.

The causal effect estimates are corrected for possible selection bias due to analysis of the patients with available HRe. One of these weights was found to be extreme, therefore we investigate the impact of its value on the results from the MSM. We gradually

truncated the extreme weight, and estimated the causal effect estimates using only  $SW_i$  weights. We observed that the more extreme the weights for observed HRe become, the larger the effects of dose reductions and cycle delays are estimated, while the effect of surgery delay decreases. However, the observed differences are quite small. For example, when all IPCWs for observed HRe are set to 1, the OR for additional delayed cycle is 2.477 ( $p$ -value 0.016), and the OR for surgery delay is 0.377 ( $p$ -value 0.019). As a result, we conclude that a small selection bias is introduced with analysis of records with known HRe.

To the extent of researchers opinion, Definition 4.18 incorporates all elements that comprise the exposure and justifies consistency (see Section 2.2.2) between factual and counter-factual outcomes. The complexity of the counter-factual outcomes generates a great amount of potential outcomes but does not falsify consistency. Lack of unmeasured confounding is required to guarantee the validity of the inference drawn from a MSM. It is an untestable assumption. In these data toxicities are hardly able to predict cycle delays and dose reductions which makes us suspicious that there might be some unmeasured confounders. Our investigation suggests that there is a great variability between oncologists and hospitals in the way decisions to modify therapy are taken, and the extent to which modifications are tolerated. This problem could have been accounted for by a random surgeon or hospital effect in the models for exposure allocation. However, information about surgeons is not available, while the variable hospital sizes do not permit the use of hospital random effect. Exchangeability builds upon lack of unmeasured confounding and as such need not be discussed when we have doubts about unmeasured confounding. One might question the validity of the positivity assumption especially concerning dose reductions, since the protocol states that once applied dose reduction should be preserved for the consecutive courses. However, there are a number of examples where full dose was given after a reduced dose cycle. This means that there are no structural zeros in the data. Further, our understanding of the reasoning behind therapy modifications permits fictitious non-compliance with the study protocol since delays and dose reductions are often used interchangeably.

This section devoted special attention to model building. This was done because the validity of the inference depends on the correct model specification assumption. To some extent a data driven approach was adopted by looking at the change in AIC value of nested models, although biological understanding backed all attempts. This was done with the aim to identify the correct model specification. As a result each component of the weights has a mean equal to one which is a property that indicates correct model specification.

As a result of the choices made in this chapter either due to the data or to the interpretation of the protocol, some caution in the interpretation of the results is required. Namely, we model the effect of small delays and small dose reductions. By studying the number of occurrences of each exposure, the analyses do not distinguish between a delay of the second and the third cycle, something that might make a difference in practice. We corrected each exposure for confounding with a subset of the recorded toxicities as they are listed in the study protocol, although all toxicities could have

been considered. This chapter presented the process of building and estimation of three MSMs for a binary outcome with increasing complexity. It showed the complexity of the processes that comprise the treatment of osteosarcoma, and how involving it is to model all interdependences in chemotherapy data. However, we are also interested in the effects of therapy modifications on patients' survival. Models for time-to-event outcomes will be discussed in the following chapter.

## Chapter 5

# MSCPHMs for survival

This chapter presents a Marginal Structural Cox Proportional Hazards Model (MSCPHM) for estimation of the causal effects of chemotherapy modifications on patients' survival. This chapter is structured as follows. First, the background of the problem is introduced. The analyses in Section 5.1 investigate the time-dependent confounding and the exposure-confounder feedback. In Section 5.2 the MSCPHM is formulated and the estimation process is described. The chapter ends by presenting and discussing the results for Event-Free-Survival and Overall Survival.

In this chapter we aim to evaluate the joint causal effects on patients' survival of the three types of therapy modifications, namely cycle and surgery delays, and dose reductions. Before conducting any type of survival analysis, the time origin must be defined. This depends on the research question. In many applications the time of randomization is chosen. However, in this study the exposures and the time-dependent confounders are not known at the time of randomization to dose-intense or control arm (see Section 3.1). Although Cox proportional hazards model can evaluate the effects of time-dependent covariates, we decide to condition survival on completion of the therapy within 180 days since therapy initiation, i.e. we perform a landmark analysis. This landmark point is relevant for clinicians. Completion of the therapy we define as undergoing surgery and all 6 chemotherapy cycles. In order to allow most of the patients to be included in the analysis while they experience therapy delays, the landmark is placed at 180 days since day 1 of cycle 1. Although according to the treatment protocol the therapy should be completed within 141 days, some patients' treatment period is longer than 180 days.

**Definition 5.1. Complete therapy** A patient has completed protocol treatment when he/she has undergone surgery and six chemotherapy cycles.

The condition to complete therapy within 180 days excludes from the analysis patients with less than 6 chemotherapy courses. Figure 3.4b shows the number of patients who do not complete therapy and the associated major reason to terminate treatment. Some of these patients could not be analysed even without the condition for completion of the therapy because of unrecorded date of surgery. The latter prevents us from calculating their exposure. In total 8 patients experience disease progression. In total 219

patients were randomized to the control arm and meet the inclusion criteria; 17 patients cannot be analysed due to unknown date of surgery. Further, 9 patients with progressive disease were excluded. Seven patients experience the event of interest before the landmark. From the analyses additional 20 patients who discontinued treatment prematurely, i.e. receive less than six cycles, were excluded. As a result, the EFS analysis is based on a sample of 161 patients, and the OS analysis on a sample of 164 patients. The number of events are 83 and 59 in the EFS and OS sample, respectively.

Figure 5.1 visualizes through a DAG all treatment patterns and causal relations observed in the analysed sample. Expert knowledge suggests that *toxicities* throughout cycle  $k$  determine the *exposure* in cycle  $k+1$ . And the *exposure* in cycle  $k+1$  is given to decrease the *toxicities* through cycle  $k+1$ . All these relationships are depicted with edges that link two nodes. *Exposure* nodes are comprised of *cycle delays* and *dose reductions*. Analogously, *toxicity* nodes denote the collection of all possible toxicities. *Surgery delay* is presented in a separate node because it depends on the number of preoperative cycles, and is observed only once per treatment trajectory. In fact it combines the previously encountered nodes of *scheduled surgery* and *surgery delay*. In this way, the node combines the two timing aspects of surgery – in cycles and in days since the end of the last preoperative cycle. Consider an example to understand this concept. Let a patient receive 3 preoperative cycles. Then the first *surgery delay* node does not exist on the subject-specific causal pathway. The same applies for the *surgery delay* vertices after cycle 4. The only *surgery delay* vertex left is the one between *toxicity cycle 3* and *exposure cycle 4*.

Toxicities and exposures are dependent in time. This is represented by edges that link consecutive covariate nodes. All patients in the analysed sample underwent surgery and six chemotherapy cycles. The whole treatment trajectory, together with the baseline covariates determine patient’s survival. This is visualised via edges that link each vertice on the graph with the outcome node – *survival*.

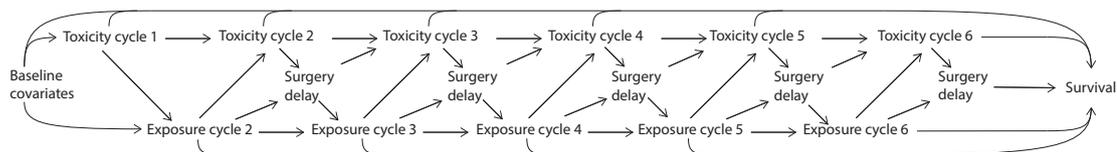


Figure 5.1: DAG describing the causal relationships between exposures, confounders and survival

## 5.1 Investigation of time-dependent confounding and exposure – confounder feedback

This section presents analyses that aim to determine which of the hypothesised relationships visualised on Figure 5.1 are present in the sample at hand and thus justify the need of a MSM to assess the joint causal effect of the exposures on the outcome.

Throughout this chapter we use the exposure and covariate definitions introduced in Chapter 4. However, this time we are modelling survival time instead of HRe, and this requires a definition of a new outcome. Definition 5.2 states the definition of EFS which we focus on in this chapter. For completeness, in the results section we also present the causal effect estimates on OS.

**Definition 5.2. EFS Outcome** Let  $T$  be a continuous variable denoting the time in days until disease progression, local recurrence, distant metastases, or death due to any cause, whichever comes first since chemotherapy initiation.

The initial step in the investigation of time-dependent confounding is to determine whether toxicities are independent risk factors for patients' survival. Since a patient could die from the treatment side-effects, we expect toxicities to affect negatively patients' survival. However, this is only true for severe toxicities. Low to moderate toxicities are considered signs of therapy effectiveness. We model toxicities in a cumulative fashion using Definitions 4.6 – 4.10 but summing from cycle 1 to cycle 6. Therefore, in this chapter we analyse the whole treatment data. The analysis of the postoperative treatment allows us to estimate the effect of cardiac toxicity in addition to leucopenia, thrombocytopenia, oral toxicity, infection and ototoxicity.

**Definition 5.3. Cumulative cardiac toxicity**  $\mathcal{L}^6 = \sum_{k=1}^6 L_k^6$  is a discrete variable denoting the number of occurrences of cardiac toxicity throughout treatment, where  $L_k^6$  denotes the occurrence of cardiac toxicity throughout cycle  $k$ , i.e

$$L_k^6 = \mathbb{1}_{\{\text{cardiac toxicity in cycle } k\}} \begin{cases} 1 & , \text{ if cardiac toxicity CTCAE grade } \geq 1 \text{ for cycle } k \\ 0 & , \text{ if cardiac toxicity CTCAE grade } = 0 \text{ for cycle } k. \end{cases} \quad (5.1)$$

We estimate the effects of the cumulative toxicities on EFS by fitting a Cox proportional hazards model. The latter are corrected for the cumulative exposures as formulated in Definitions 4.11 and 4.12 but summing over all six cycles, surgery delay according to Definition 4.17, and baseline covariates – age and gender. This model is expressed as follows:

$$\lambda_T(t \mid \mathcal{L}, \mathcal{A}, \mathbf{V}) = \lambda_0(t) \exp(\delta_1 \mathcal{L}^1 + \delta_2 \mathcal{L}^2 + \delta_3 \mathcal{L}^3 + \delta_4 \mathcal{L}^4 + \delta_5 \mathcal{L}^5 + \delta_6 \mathcal{L}^6 + \delta_7 \mathcal{A}^1 + \delta_8 \mathcal{A}^2 + \delta_9 \mathcal{A}^3 + \delta_{10} \mathbb{1}_{\{\text{female}\}} + \delta_{11} \text{age}), \quad (5.2)$$

where  $\lambda_T(t)$  is the hazard rate function, which gives the rate at which an individual, who survived to time  $t$ , will experience the event in the next instant of time, and  $\lambda_0(t)$  denotes the baseline hazard.

The  $\delta$ -parameters in Model (5.2) estimate the logarithm of the hazard ratio for an additional cycle with toxicity (exposure) vs. no additional cycle with toxicity (exposure). These estimates are listed in Table 5.1. None of the toxicities has an effect significantly different from zero. This could be due to incorrect model specification, or because EFS

events do not occur short after chemotherapy to be affected by toxicities, although one EFS event occurred on day 23 since completion of the therapy. We found no association between cycle delays and dose reductions and the hazard of an EFS event. On the contrary, surgery delay increases the hazard of an EFS event.

Table 5.1: Toxicities effect estimates on EFS

	Estimate*	SE(Estimate)	<i>p</i> -value
Cumulative leucopenia	-0.005	0.081	0.950
Cumulative thrombocitopenia	-0.008	0.132	0.952
Cumulative mucositis	-0.033	0.093	0.718
Cumulative infection	0.067	0.073	0.363
Cumulative ototoxicity	0.010	0.192	0.959
Cumulative cardiac toxicity	-0.229	0.233	0.325
Cumulative cycle delay	0.024	0.087	0.788
Cumulative dose reduction	-0.064	0.065	0.318
Surgery delay	0.337	0.229	0.141
Female	-0.400	0.239	0.094
Age	0.029	0.018	0.111

\* log hazard ratio

Next we use the toxicity data from the first five cycles and the exposure data from cycle 2 onwards to evaluate the effect of toxicities on exposure allocation. According to the treatment protocol cycle  $k$  is delayed if myelosuppression or mucositis occurs during cycle  $k - 1$ . Allocation of cycle delay is modelled as a dichotomous variable and therefore we build a pooled logistic regression model to evaluate the effect of toxicities on the exposure. Such a logistic regression model treats every patient-cycle as an independent observation. Pooled logistic regression allows us to examine toxicity-exposure relationship in a cross-sectional fashion while still preserving the longitudinal nature of the processes. The exposure in the previous cycle is supposed to decrease the toxicities throughout the cycle, which then determine the exposure in the next cycle. Therefore, we need to correct the effect of toxicities throughout cycle  $k - 1$  on the exposure in cycle  $k$  for the exposure in cycle  $k - 1$ . However, if we do not in addition correct for the baseline covariates, we might allow an indirect effect of toxicities on exposure to be transmitted through the baseline confounders. We therefore estimate model

$$P(A_k^1 = 1 \mid L_{k-1}^1, L_{k-1}^2, L_{k-1}^3, A_{k-1}^1, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 L_{k-1}^1 + \delta_2 L_{k-1}^2 + \delta_3 L_{k-1}^3 + \delta_4 A_{k-1}^1 + \delta_5 \mathbb{1}_{\{\text{female}\}} + \delta_6 \text{age}), \quad (5.3)$$

for  $k = 2, 3, \dots, 6$ , where  $A_1^1$  is set to zero for all patients.

The log-odds estimates of Model (5.3) are listed in Table 5.2. Both leucopenia and delay of the previous cycle increase the probability of delay of the current cycle. These findings meet researchers' expectations. Corrected for the rest, occurrence of mucositis

also has a positive effect on cycle delays, although the evidence for this relationship is not as strong. In presence of leucopenia, thrombocytopenia does not add extra information to the model (the estimated effect is practically equal to 0). This is a logical result since both leucopenia and thrombocytopenia indirectly measure myelosuppression, i.e. are expected to be highly collinear.

Table 5.2: Toxicities effect estimates on administration of delays

	Estimate*	SE(Estimate)	<i>p</i> -value
(Intercept)	-0.927	0.128	0.000
Leucopenia	0.749	0.173	0.000
Thrombocytopenia	-0.060	0.298	0.840
Mucositis	0.381	0.210	0.070
Delay of cycle $k-1$	1.016	0.163	0.000
Female	-0.100	0.155	0.519
Age	-0.002	0.012	0.864

\* log-odds

In Chapter 4 we defined variables  $F_k^1$  and  $F_k^2$  which indicate a joint occurrence of myelosuppression (expressed as leucopenia and thrombocytopenia) and infection to model allocation of dose reductions. Here we augment the set of  $F$ -variables to account for severe oral toxicity that also justifies the application of dose reductions. It was not possible to use the latter in the models in Chapter 4 because severe toxicities occur mainly during postoperative chemotherapy cycles.

**Definition 5.4. Severe oral toxicity**  $F_k^3$  is a dichotomous variable denoting occurrence of severe oral toxicity throughout cycle  $k$ , i.e.

$$F_k^3 = \mathbb{1}_{\{\text{severe mucositis in cycle } k\}} \begin{cases} 1 & , \text{ if oral toxicity CTCAE grade } \geq 3 \text{ for cycle } k \\ 0 & , \text{ if oral toxicity CTCAE grade } < 3 \text{ for cycle } k. \end{cases} \quad (5.4)$$

The treatment protocol further suggests discontinuation of CDDP and DOX in case of ototoxicity and cardiac toxicity, respectively. Similar to Model (5.3), we model the effect of toxicities on allocation of dose reduction by fitting a pooled logistic regression model, and correct these estimates for the exposure in the previous cycle and baseline covariates. This means estimating model

$$P(A_k^2 = 1 \mid F_{k-1}^1, F_{k-1}^2, F_{k-1}^3, L_{k-1}^5, L_{k-1}^6, A_{k-1}^2, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 F_{k-1}^1 + \delta_2 F_{k-1}^2 + \delta_3 F_{k-1}^3 + \delta_4 L_{k-1}^5 + \delta_5 L_{k-1}^6 + \delta_6 A_{k-1}^2 + \delta_7 \mathbb{1}_{\{\text{female}\}} + \delta_8 \text{age}), \quad (5.5)$$

where  $k$  takes values from 2 to 6, and  $A_1^2$  is set to zero for all patients.

Model (5.5) parameter estimates are listed in Table 5.3. All toxicities increase the probability of allocation of dose reduction. This finding is in line with researchers'

expectations. Severe oral and ototoxicity have large effect estimates even when corrected for all the rest. The dose of older patients is more often reduced, while there are no gender differences. The results in Table 5.3 also confirm that once applied dose reductions are preserved for subsequent cycles by the very large positive effect of dose reduction in the previous cycle.

Table 5.3: Toxicities effect estimates on administration of dose reductions

	Estimate*	SE(Estimate)	<i>p</i> -value
(Intercept)	-2.166	0.177	0.000
Leucopenia for dose reduction	0.327	0.460	0.477
Thrombocytopenia for dose reduction	0.732	0.634	0.249
Severe oral toxicity	1.080	0.432	0.013
Ototoxicity	1.001	0.627	0.111
Cardiac toxicity	0.323	0.798	0.686
Dose reduction in cycle $k-1$	5.055	0.400	0.000
Female	-0.031	0.248	0.902
Age	0.064	0.018	0.000

\* log-odds

Surgery delay is another form of delay of subsequent treatment and as such is modelled in the same fashion as cycle delays. The only exception is that surgery delay is observed only once, and hence we fit an ordinary logistic regression in contrast to the pooled logistic regressions used for cycle delays and dose reductions. The toxicities from the last preoperative cycle of each patient are used to predict surgery delay. Their estimates are corrected for baseline covariates. The model for allocation of surgery delay is formulated as follows:

$$P(A^3 = 1 \mid L_s^1, L_s^2, L_s^3, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 L_s^1 + \delta_2 L_s^2 + \delta_3 L_s^3 + \delta_4 \mathbb{1}_{\{\text{female}\}} + \delta_5 \text{age}), \quad (5.6)$$

where  $s$  denotes the last preoperative cycle.

Parameter estimates of Model (5.6) are presented in Table 5.4. These results are in line with the observed fact that many surgeries were delayed due to administrative reasons rather than patient's condition. In the analysed sample rather patients with surgery on time experience toxicities during the last preoperative cycle, than those with a delayed surgery. However, none of the effect estimates is significantly different from zero. These results are comparable with those presented in Table 4.14 in Section 4.3.1. Patient's age and gender are also not associated with allocation of surgery delays. Only 15.8% of the surgeries were delayed due to haematological toxicity. This is likely the reason why we observe these results.

The models for exposure allocation confirm that at least partially therapy modifications are driven by patients' toxicities. We continue our investigation of the exposure-confounder feedback by modelling toxicities as a function of the toxicities in the previous

Table 5.4: Toxicities effect estimates on allocation of surgery delay

	Estimate*	SE(Estimate)	<i>p</i> -value
(Intercept)	-0.253	0.249	0.312
Leucopenia	-0.301	0.438	0.494
Thrombocytopenia	-0.111	0.799	0.890
Mucositis	-0.950	0.558	0.091
Female	0.049	0.344	0.886
Age	0.033	0.027	0.225

\* log-odds

cycle and the exposure in the current cycle. Therapy modifications are meant to decrease subsequent toxicity or prevent from its subsequent occurrence. We expect negative effect estimates of exposures on occurrence of side effects. We use chemotherapy data from cycles 2 through 6, and build pooled logistic regressions to model the occurrences of toxicities.

Cycle delay and dose reduction are induced by occurrences of leucopenia, thrombocytopenia, and oral toxicity. Therefore these therapy modifications should decrease the probability of subsequent occurrence of these toxicities. We build the same model for these three side effects and try to explain their occurrence by current cycle delay and dose reduction. In these models we correct for the set of six toxicities from the previous cycle and baseline covariates. The general form of these models is:

$$P(L_k^j = 1 \mid \mathbf{A}_k, \mathbf{L}_{k-1}, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 A_k^1 + \delta_2 A_k^2 + \boldsymbol{\delta}_3^\top \mathbf{L}_{k-1} + \delta_4 \mathbb{1}_{\{\text{female}\}} + \delta_5 \text{age}), \quad (5.7)$$

where  $j = 1, 2, 3$ ,  $k = 2, 3, \dots, 6$  and  $\mathbf{L}_{k-1}$  is the collection of all toxicities experienced in cycle  $k - 1$ , i.e.  $\mathbf{L}_{k-1} = (L_{k-1}^1, L_{k-1}^2, L_{k-1}^3, L_{k-1}^4, L_{k-1}^5, L_{k-1}^6)$  as defined in Chapter 4.

In the same fashion, infection, oto- and cardiac toxicity justify allocation of dose reductions. Therefore, dose reduction must be given in order to prevent subsequent occurrence of these side-effects. We model these relationships using current cycle dose reduction, toxicities from the previous cycle and baseline covariates. We formulate each of these models for  $j = 4, 5, 6$  as:

$$P(L_k^j = 1 \mid A_k^2, \mathbf{L}_{k-1}, \mathbf{V}) = \text{expit}(\delta_0 + \delta_1 A_k^2 + \boldsymbol{\delta}_2^\top \mathbf{L}_{k-1} + \delta_3 \mathbb{1}_{\{\text{female}\}} + \delta_4 \text{age}), \quad (5.8)$$

where  $k = 2, 3, \dots, 6$ . Note that both Models (5.7) and (5.8) are pooled-logistic regression models where each patient-cycle is treated as an independent observation.

In the interest of presentation, Table 5.5 lists only the effect estimates of the exposures on the occurrence of toxicities as defined in Models (5.7) and (5.8). Unexpectedly, the exposures increase the probability of leucopenia of grade 2 or larger. Similar to the

lack of association between thrombocytopenia and subsequent exposure, the effects of dose reductions and cycle delays on occurrence of thrombocytopenia are not statistically significantly different from zero. On the contrary, both cycle delays and dose reductions decrease the probability of subsequent oral toxicity of grade 2 or severer. Further, dose reductions alone have a negative effect on occurrence of infection and a strong positive effect on cardiac toxicity. This discrepancy can be attributed to the following. Infection as well as oral toxicity is a clear predictor of dose reductions (see Table 5.3). At the same time, dose reduction prevents subsequent occurrence of these side effects. However, this sample does not provide indication that dose reduction has the same effect on cardiac toxicity. At the same time, cardiac toxicity occurs only in late cycles when doses are more frequently reduced. This results in a strong positive association between dose reductions and cardiac toxicity. The dependence between current cycle dose reduction and development of ototoxicity is less pronounced than the dependence between ototoxicity and dose reduction during the following cycle (see Table 5.3). The observed positive association can be explained in the same way as we did for dose reductions and cardiac toxicity. Dose reduction also has a strong negative effect (-1.208 (0.469)) on subsequent occurrence of severe oral toxicity (result not shown).

Table 5.5: Effect estimates of exposure on subsequent occurrence of toxicities

Effect of	Effect on	Estimate*	SE(Estimate)	<i>p</i> -value
Cycle delay	Leucopenia	0.408	0.162	0.012
Dose reduction	Leucopenia	0.224	0.179	0.211
Cycle delay	Thrombocytopenia	0.144	0.231	0.534
Dose reduction	Thrombocytopenia	-0.186	0.264	0.481
Cycle delay	Oral toxicity	-0.582	0.233	0.013
Dose reduction	Oral toxicity	-0.660	0.264	0.012
Dose reduction	Infection	-0.417	0.236	0.077
Dose reduction	Ototoxicity	0.122	0.422	0.772
Dose reduction	Cardiac toxicity	1.200	0.458	0.009

\* log-odds

All results are based on a number of compromises that we had to make. Due to lack of instances of exposure otherwise, cycle delays are predominantly very short, and reductions range from 14 to 100% (discontinuation). These limitations could explain both the lack of association where expected, and the presence of associations with counter-intuitive sign.

## 5.2 Model formulation and estimation

In the previous section we showed that the hypothesised exposure-confounder feedback can be identified in the sample at hand. Therefore, in order to assess the joint causal effect of surgery and cycle delays, and dose reductions on patients' survival we build a MSM. As stated in the beginning of the chapter, we discuss in detail the analysis for EFS

but the results for both EFS and OS will be presented. MSMs model the distribution of counter-factual outcomes.

**Definition 5.5. Counter-factual EFS outcome** Let  $T_{(\bar{a}^1, \bar{a}^2, a^3)}$  be the EFS time arising from a six-cycle chemotherapy where  $A_k^j = a_k^j$ , and a surgery delay  $A^3 = a^3$ , where  $a_k^j$  and  $a^3 \in \{0, 1\}$ , for  $j = 1, 2$ , and  $k = 2, 3, \dots, 6$ , i.e.  $(\bar{a}^1, \bar{a}^2, a^3) = ((a_2^1, a_2^2), (a_3^1, a_3^2), (a_4^1, a_4^2), (a_5^1, a_5^2), (a_6^1, a_6^2), a^3)$ .

From Definition 5.5 follows that there are 2048 ( $2^{11}$ ) counter-factual outcomes arising from all combinations of treatment trajectories defined by 11 binary variables. The estimation of their distribution is a challenging problem when we observe the records of hardly 161 patients. In order to estimate counter-factual outcomes outside the domain of the observed outcomes, the model extrapolates the dependences between the observed outcomes. A MSCPHM for the joint effect of therapy modifications on EFS is defined as:

$$\lambda_{T_{(\bar{a}^1, \bar{a}^2, a^3)}}(t) = \lambda_0(t) \exp(\beta_1 \mathcal{A}^1 + \beta_2 \mathcal{A}^2 + \beta_3 A^3), \quad (5.9)$$

where  $\mathcal{A}^1$  and  $\mathcal{A}^2$  denote cumulative cycle delays and dose reductions as formulated in Definitions 4.11 and 4.12, respectively, but summing over the whole treatment trajectory.

Model (5.9) is a MSM for the counter-factual EFS outcome according to Definition 5.5. However, we observe only a single outcome per patient because we cannot apply multiple treatment strategies and observe multiple survival times for the same patient. This prevents us from directly estimating the parameters of Model (5.9) using the available data. Instead, we fit a weighted Cox regression to model the factual outcomes of the patients. Such a model is defined as:

$$\lambda_T(t \mid \mathcal{A}, A^3) = \lambda_0(t) \exp(\gamma_1 \mathcal{A}^1 + \gamma_2 \mathcal{A}^2 + \gamma_3 A^3), \quad (5.10)$$

where  $T$  is the survival time according to Definition 5.2.

Model (5.10) is the factual analogue of Model (5.9). In order to obtain unbiased estimates of the  $\beta$ -parameters of the MSM, we should fit Model (5.10) in a population where exposure allocation is independent of side-effects rather than the study sample. The pseudo-population is created by weighting the observations by the inverse of the probability of observed exposure. In this analysis there are three exposures. This means that we should weight patients' treatment trajectories by the joint probability distribution of all three exposures throughout the six cycles. However, as shown earlier in this thesis, there is a way to simplify this task by using conditional probability rule. For the application of this rule it is very important to take into account the temporal order of occurrence of the exposures. At the beginning of each cycle, the cycle delay is observed. Then the dose to be administered throughout the first three days of each cycle is determined. When a cycle starts, we observe whether dose reduction took place. Therefore, when modelling administration of dose reductions, we can condition on current cycle delay. At the last preoperative cycle we have observed all preoperative cycle delays and dose reductions. That is why we can afford to condition surgery delay on those.

By applying the property  $P(A, B, C) = P(A | B, C) \times P(B | C) \times P(C)$ , the joint probability of the three exposures can be easily estimated. Here  $A$  represents surgery delay,  $B$  dose reduction, and  $C$  cycle delay. In addition two of three exposures are longitudinal. We can look at the distribution of six cycle delays as a joint distribution, and apply the same rule for decomposition of joint probability. However, this is not the best approach in this case. Alternatively, we can construct independent events, and compute the joint probability as a product of independent events. To construct independent events we model cycle-wise the occurrence of cycle delay and dose reduction by conditioning them on the exposure from the preceding cycle. This is only possible because the dependence between the exposures that are two or more cycles apart, is transmitted by the exposure from the preceding cycle. This means that conditional on cycle  $k-1$  delay, the delays of cycle  $k$  and  $k-2$  are assumed independent. These relations are expressed as:

$$\begin{aligned}
P(\bar{A}^1, \bar{A}^2, A^3) &= P(\bar{A}^1) \times P(\bar{A}^2 | \bar{A}^1) \times P(A^3 | \bar{A}^1, \bar{A}^2) \\
&= \prod_{k=2}^6 [P(A_k^1 | A_{k-1}^1, A_{k-1}^2) \times P(A_k^2 | A_k^1, A_{k-1}^1, A_{k-1}^2)] \times P(A^3 | A_s^1, A_s^2),
\end{aligned}
\tag{5.11}$$

where  $\bar{A}^j = (A_2^j, A_3^j, A_4^j, A_5^j, A_6^j)$ , for  $j = 1, 2$ , and  $s$  denotes the number of preoperative cycles.

Here we make two assumptions: conditioning only on the previous cycle exposure is enough to break the temporal dependence between the exposures. This is a standard assumption in applications of MSMs (e.g. [19, 20]). Caution is required when this assumption is applied in another field of study. The second assumption claims that surgery delay depends only on the exposure during the last preoperative cycle.

We start with a model for allocation of cycle delays as a function of patient's toxicities and baseline covariates to estimate the first term of the product on the right-hand side of Equation (5.11).

In this chapter we devote special attention to the process of building correct models for exposure allocation because the validity of the inference through MSMs depends on this assumption. Since there are five instances of cycle delay allocation, which is modelled as a binary variable, we build a pooled logistic regression where each patient-cycle is considered as an independent observation. These models have three important elements: incorporation of time into the model, possible non-linear function of patient's age, and interaction of age and gender.

The incorporation of time into the pooled logistic regression model for allocation of cycle delays could be done in one of the following ways: by using a linear, quadratic, or a higher order function of time, or model cycles independently with an indicator function for each cycle. The latter is possible because there are only five time instances. By considering linear function of time as a benchmark, none of the explored non-linear functions led to a better model fit, as measured by AIC. However, using independent

cycle indicators increased model's goodness-of-fit (likelihood ratio test  $p$ -value is 0.048). Using this time specification, linear, quadratic, and RCS (with 3 knots) functions of age were tested but none of them significantly enhanced model's fit.

As a third step, we added interaction terms between different functions of age and gender but that decreased the AIC value very slightly. A selected list of all tested model formulations is shown in Table 5.6. Although specification 3 has only one unit lower AIC value compared to specification 2 while it includes 5 additional parameters, we decided to keep it in order to provide a flexible model. As a result the effect of cycle delays on EFS almost doubles. We also decided to keep the interaction between RCS function of age and gender. This decision is driven mainly by the understanding that the biological clock of women differs from the one of men. As a result, using specification 4 we increase the variability of the weights, which is very low anyway, and achieve minimum AIC. This model results in average weight of 1, which is regarded as an indication of correct model specification [5]. The effect of cycle delay on EFS changes in the direction towards its estimate when simpler model is used. The mathematical formulation of specification 4 is as follows:

$$P(A_k^1 = 1 | L_{k-1}^1, L_{k-1}^2, L_{k-1}^3, \mathbf{A}_{k-1}, \mathbf{V}) = \text{expit}(\alpha_0 + \alpha_1 \mathbb{1}_{\{\text{third cycle}\}} + \alpha_2 \mathbb{1}_{\{\text{fourth cycle}\}} + \alpha_3 \mathbb{1}_{\{\text{fifth cycle}\}} + \alpha_4 \mathbb{1}_{\{\text{sixth cycle}\}} + \alpha_5 L_{k-1}^1 + \alpha_6 L_{k-1}^2 + \alpha_7 L_{k-1}^3 + \alpha_8 A_{k-1}^1 + \alpha_9 A_{k-1}^2 + \alpha_{10}^\top RCS(\text{age}, 3) \times \mathbb{1}_{\{\text{female}\}}), \quad (5.12)$$

for  $k = 2, \dots, 6$ , where  $RCS(\text{age}, 3)$  denotes RCS function with 3 knots, and  $\alpha_0$  is the log-odds of cycle 2 delay for a 16 years old male patient with no toxicity throughout cycle 1.

The estimates of the logarithm of hazard ratio (last column in Table 5.6) are obtained by fitting Model (5.10) to a pseudo-population created by stabilised weights under the respective exposure allocation model specification. The stabilised weights are the ratio of the fitted probabilities of Model (5.13) defined as:

$$P(A_k^1 = 1 | \mathbf{A}_{k-1}) = \text{expit}(\alpha_0 + \alpha_1 \mathbb{1}_{\{\text{third cycle}\}} + \alpha_2 \mathbb{1}_{\{\text{fourth cycle}\}} + \alpha_3 \mathbb{1}_{\{\text{fifth cycle}\}} + \alpha_4 \mathbb{1}_{\{\text{sixth cycle}\}} + \alpha_5 A_{k-1}^1 + \alpha_6 A_{k-1}^2), \quad (5.13)$$

and Model (5.12), evaluated at the observed exposure trajectory for each patient. Models (5.12) and (5.13) estimate the probability of cycle delay conditional on treatment and confounder history and treatment history only, respectively. Let us denote these probabilities by  $p_{i,k}^{del}$  and  $p_{i,k}^{del,*}$ , respectively. The stabilised weights are given as:

$$SW_i^{del} = \prod_{k=2}^6 \frac{[(p_{i,k}^{del,*})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{del,*})^{1-A_{i,k}^1}]}{[(p_{i,k}^{del})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{del})^{1-A_{i,k}^1}]}. \quad (5.14)$$

Table 5.6: Models for allocation of cycle delays using all treatment data

	Specification	AIC <sup>†</sup>	Mean weight (SD <sup>†</sup> )	Min/Max	Estimate* (SE <sup>†</sup> )
1	Linear functions of cycle and age, indicators of previous cycle delay and dose reduction, leucopenia, thrombocytopenia, mucositis, and gender	1029.3	1.000 (0.376)	0.270/2.992	-0.051 (0.084)
2	As per model 1 but cycle number indicators instead of a linear function	1027.4	0.998 (0.372)	0.276/2.835	-0.049 (0.085)
3	As per model 2 but 3-knot RCS <sup>†</sup> for age instead of a linear function	1026.4	1.003 (0.422)	0.268/2.737	-0.085 (0.086)
4	As per model 3 but including interaction between age and gender	1025.4	1.000 (0.448)	0.243/3.180	-0.077 (0.086)

\* log hazard ratio

† AIC, Akaike's information criterion; SD, standard deviation; SE, standard error; RCS, restricted cubic spline.

Table 5.7: Models for allocation of dose reduction using all treatment data

	Specification	AIC <sup>†</sup>	Mean weight (SD <sup>†</sup> )	Min/Max	Estimate* (SE <sup>†</sup> )
1	Linear functions of cycle and age, indicators of previous cycle delay and dose reduction, current cycle delay, leucopenia, thrombocytopenia, and mucositis for dose reduction, cardiac and ototoxicity, and gender	522.2	0.978 (0.369)	0.239/2.556	-0.072 (0.059)
2	As per model 1 but cycle number indicators instead of a linear function	517.7	0.979 (0.378)	0.189/2.709	-0.073 (0.059)
3	As per model 2 but using separate indicators for myelosuppression and infection	497.1	1.006 (0.733)	0.171/5.412	-0.187 (0.067)
4	As per model 3 but including interaction terms between leucopenia and infection, and thrombocytopenia and infection	501.1	1.008 (0.744)	0.166/5.575	-0.188 (0.067)

\* log hazard ratio

† AIC, Akaike's information criterion; SD, standard deviation; SE, standard error.

The numerator and the denominator of Equation (5.14) resemble binomial likelihood. This functional form results in the probability of observed exposure at each cycle for each patient. By taking the product over cycles 2 through 6, the probability of observed exposure for the whole treatment trajectory is computed. As such, the numerator evaluates the probability of the observed exposure given exposure history only, and the denominator the probability of the observed exposure given exposure and covariate history.

The model for allocation of dose reductions has one more aspect in addition to time: age and gender which should be carefully specified. Many more toxicities predict dose reductions and some of them are combined in joint indicators as prescribed by the treatment protocol (the  $F$  variables from Chapter 4). This leads to many possibilities for modelling toxicities.

Similar to the model for cycle delays, in the model for dose reductions independent cycle indicators significantly improve the goodness-of-fit of the model. This modification and its implications are presented under specification 2 in Table 5.7. Further, we tested different non-linear functions of age and possible interactions with gender but none of them improved model's goodness-of-fit as measured by AIC. Separating the joint indicators of leucopenia and infection, and thrombocytopenia and infection resulted in much better fit. This model is described in specification 3. We further considered reasonable to include interaction terms between the two forms of myelosuppression and infection because leucopenia alone is supposed to trigger cycle delay, while leucopenia together with infection should bring to dose reduction. This modification did not change the results (the effect of dose reduction on EFS increases with 0.001 on the log-odds scale).

Although the effect of dose reductions on EFS more than doubled when using model specification 3 compared to specifications 1 and 2, the weights reached a mean closer to 1, which is an indication of correct model specification. Model 3 in Table 5.7 is expressed as:

$$\begin{aligned}
P(A_k^2 = 1 \mid L_{k-1}^1, L_{k-1}^2, F_{k-1}^3, L_{k-1}^4, L_{k-1}^5, L_{k-1}^6, A_k^1, \mathbf{A}_{k-1}, \mathbf{V}) = \text{expit}(\alpha_0 \\
+ \alpha_1 \mathbb{1}_{\{\text{third cycle}\}} + \alpha_2 \mathbb{1}_{\{\text{fourth cycle}\}} + \alpha_3 \mathbb{1}_{\{\text{fifth cycle}\}} + \alpha_4 \mathbb{1}_{\{\text{sixth cycle}\}} + \alpha_5 L_{k-1}^1 + \alpha_6 L_{k-1}^2 \\
+ \alpha_7 F_{k-1}^3 + \alpha_8 L_{k-1}^4 + \alpha_9 L_{k-1}^5 + \alpha_{10} L_{k-1}^6 + \alpha_{11} A_{k-1}^1 + \alpha_{12} A_{k-1}^2 + \alpha_{13} A_k^1 + \alpha_{14} \text{age} \\
+ \alpha_{15} \mathbb{1}_{\{\text{female}\}}). \quad (5.15)
\end{aligned}$$

In order to calculate stabilised weights we additionally fit model

$$\begin{aligned}
P(A_k^2 = 1 \mid A_k^1, \mathbf{A}_{k-1}) = \text{expit}(\alpha_0 + \alpha_1 \mathbb{1}_{\{\text{third cycle}\}} + \alpha_2 \mathbb{1}_{\{\text{fourth cycle}\}} + \alpha_3 \mathbb{1}_{\{\text{fifth cycle}\}} \\
+ \alpha_4 \mathbb{1}_{\{\text{sixth cycle}\}} + \alpha_5 A_{k-1}^1 + \alpha_6 A_{k-1}^2 + \alpha_7 A_k^1). \quad (5.16)
\end{aligned}$$

The latter is used to estimate the subject- and cycle-specific probability of dose reduction given exposure history denoted by  $p_{i,k}^{red,*}$ . We use Model (5.15) to estimate the subject- and cycle-specific probability of dose reduction given exposure and covariate history

denoted by  $p_{i,k}^{red}$ . Those probabilities, together with the observe exposure trajectory are plugged in equation

$$SW_i^{red} = \prod_{k=2}^6 \frac{[(p_{i,k}^{red,*})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{red,*})^{1-A_{i,k}^1}]}{[(p_{i,k}^{red})^{A_{i,k}^1}] \times [(1 - p_{i,k}^{red})^{1-A_{i,k}^1}]} \quad (5.17)$$

to calculate subject-specific weights. The effect estimates in Table 5.7 were produced by weighting each patient's record by the product  $SW_i^{del} \times SW_i^{red}$ , where  $SW_i^{del}$  is calculated using Models (5.12) and (5.13).

We followed the same model selection procedure to identify the model for allocation of surgery delay. Since surgery takes place mid-way in the therapy, we use only the preoperative treatment trajectory. More specifically, since we make the assumption that only last exposure and toxicity influences surgery delay, we use only the information from the last preoperative cycle of each patient. As a result, we model the probability of surgery delay using simple logistic regression. We found that a linear function of time, which indicates how many preoperative cycles were given, provides the best fit. We use leucopenia, thrombocytopenia, and oral toxicity from the last preoperative cycle to predict surgery delay, and correct their effects for cycle delay and dose reduction from the last preoperative cycle, and baseline covariates. In addition, we included a quadratic function of patient's age without interaction with gender. The resulted model produces weights with a mean equal to 0.991 (0.215) and range between 0.361 and 1.927. The model is defined as:

$$P(A^3 = 1 \mid s, L_s^1, L_s^2, L_s^3, A_s^1, A_s^2, \mathbf{V}) = \text{expit}(\alpha_0 + \alpha_1 s + \alpha_2 L_s^1 + \alpha_3 L_s^2 + \alpha_4 L_s^3 + \alpha_5 A_s^1 + \alpha_6 A_s^2 + \alpha_7 \text{age} + \alpha_8 \text{age}^2 + \alpha_9 \mathbb{1}_{\{\text{female}\}}), \quad (5.18)$$

where  $s$  stands for the number of preoperative cycles.

Similar to the other elements of the exposure, we also estimated a model for surgery delay based on exposure history only. That is fitting model

$$P(A^3 = 1 \mid s, A_s^1, A_s^2) = \text{expit}(\alpha_0 + \alpha_1 s + \alpha_2 A_s^1 + \alpha_3 A_s^2). \quad (5.19)$$

We denote the probabilities estimated from Models (5.19) and (5.18) by  $p_i^{srg,*}$  and  $p_i^{srg}$ , respectively. Then the corresponding stabilised weights are calculated as

$$SW_i^{srg} = \frac{[(p_i^{srg,*})^{A_i^3}] \times [(1 - p_i^{srg,*})^{1-A_i^3}]}{[(p_i^{srg})^{A_i^3}] \times [(1 - p_i^{srg})^{1-A_i^3}]} \quad (5.20)$$

Since patients are operated only once, the surgery component of the weights is a ratio of two terms only.

The final subject-specific weight denoted by  $SW_i$  is a product of the three quantities expressed in Equations (5.14), (5.17), and (5.20), i.e.

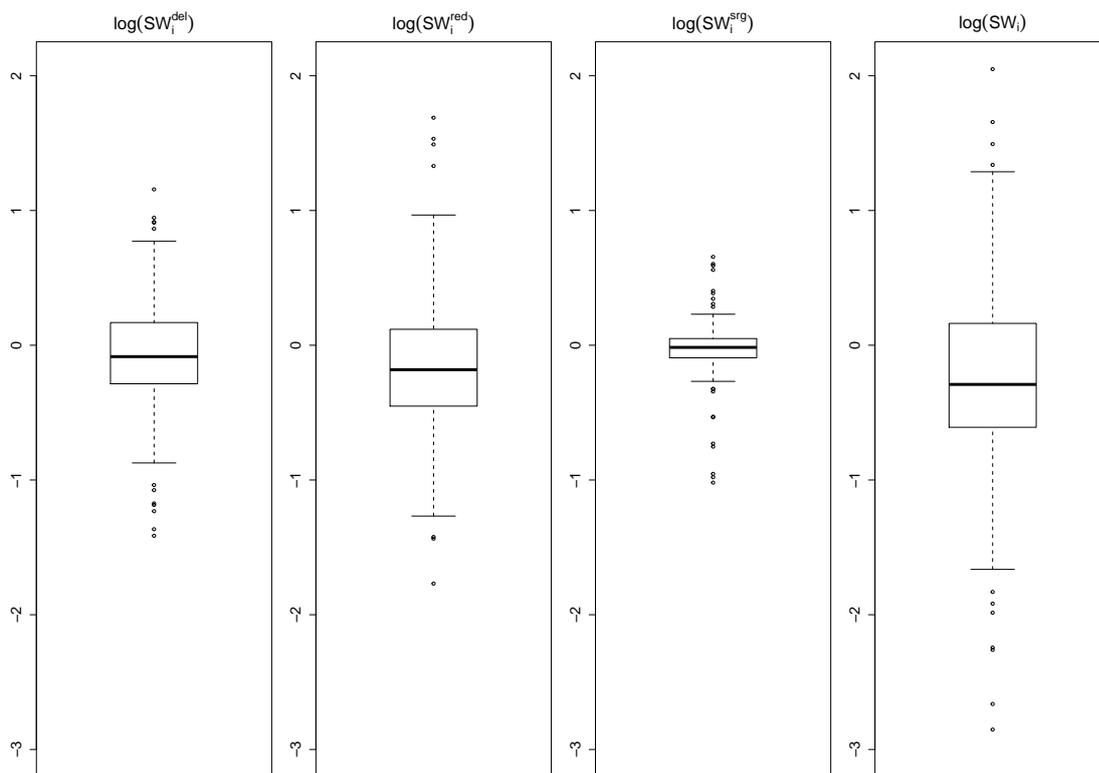


Figure 5.2: Boxplots of the final weights and all its components

$$SW_i = SW_i^{del} \times SW_i^{red} \times SW_i^{srg}. \quad (5.21)$$

The distribution of each component of the weights, as well as the final weights are visualised through boxplots on the logarithmic scale in Figure 5.2. Logarithmic scale is used since the weights have a lower bound of 0 while they can extend to infinity. The weights for cycle delays are symmetrically distributed around 1 (0 on the logarithmic scale) with very few extreme values. The same holds for dose reductions except that their range is larger compared to the weights for cycle delay. The distribution of surgery delay component of the weights is much more peaked around the value of 1. The final weights range between 0.058 and 7.761, and their mean is equal to 0.996. More than half of the patients are estimated weights smaller than 1 (the median visualised on the boxplot is located below the value of 0 on the logarithmic scale). This means that most of the patients represent examples of the hypothesised relationships between toxicities and exposures. About 30% of the patients were estimated positive log-weights and are used to counter-balance the association between exposures and toxicities.

### 5.3 Results and discussion

In this section the causal and association (results from unweighed analysis) parameter estimates for the joint effect of the three exposures on EFS and OS are presented. Table 5.8 shows the effect estimates on EFS. Each additional cycle delay causes on average a 10% increase in the hazard of an adverse event. However, this effect is not statistically significant. The corresponding association parameter indicates only a 1.3% increase in the hazard. On the contrary, an extra course with reduced dose causes a 19.5% decrease in the hazard of recurrence, progression, metastases or death. The estimate of the causal effect is statistically significant at 5% significance level. At the same time, another chemotherapy course with reduced dose is associated with only a 5% decrease in the hazard of an event. As shown in Table 5.8 both hazard ratios for surgery delay indicate a strong negative effect on the event. However, there is not enough evidence to confirm this association.

Table 5.8: Causal and association hazard-ratio estimates for the joint effect of cycle delays, dose reductions, and surgery delay on event-free survival

	HR	95% CI	<i>p</i> -value
Causal: cycle delay	1.103	[0.941 : 1.294]	0.227
Causal: dose reduction	0.837	[0.730 : 0.960]	0.011
Causal: surgery delay	1.303	[0.824 : 2.060]	0.258
Association: cycle delay	1.013	[0.862 : 1.189]	0.878
Association: dose reduction	0.953	[0.848 : 1.072]	0.426
Association: surgery delay	1.422	[0.916 : 2.206]	0.116

To estimate the joint causal effect of the therapy modifications on OS we use the same MSM and models for exposure allocation. We did not reconsider the model building for the weights because they do not depend on the outcome, and the samples are almost identical. The final weights range between 0.050 and 5.938, and have a mean equal to 0.988. Similar to EFS, an additional delayed cycle is expected to cause on average a 11.6% increase in the occurrence of event. The association model did not find any evidence for a relation between cycle delays and OS. Each additional cycle with reduced dose is estimated to cause on average a 17.5% decrease in the hazard of death. Yet, no association was found by the unweighed analysis.

Contrary to the findings in the EFS analysis, surgery delay is estimated to cause an average decrease of 13.4% in the hazard of death. The corresponding association effect estimate is smaller but has the same sign. Nevertheless, the surgery delay effect estimates on OS are far from significant. There is no reason to believe that surgery delay could have a positive effect on EFS and a negative on OS. If true, then this would mean that surgery delay has a protective effect on death but increases the risk of recurrence, progression, or metastases. However, the effects on OS are within the confidence intervals for the effect on EFS. The reverse is also true. Therefore, these results are not contradicting each other. These results are further elaborated on in Chapter 6.

Table 5.9: Causal and association hazard-ratio estimates for the joint effect of cycle delays, dose reductions, and surgery delay on overall survival

	HR	95% CI	<i>p</i> -value
Causal: cycle delay	1.116	[0.922 : 1.351]	0.260
Causal: dose reduction	0.851	[0.718 : 1.010]	0.065
Causal: surgery delay	0.882	[0.494 : 1.575]	0.671
Association: cycle delay	1.003	[0.833 : 1.207]	0.976
Association: dose reduction	0.982	[0.854 : 1.130]	0.804
Association: surgery delay	0.931	[0.544 : 1.594]	0.795

The analyses in this section show how to build a MSCPHM to estimate the joint causal effects of surgery and cycle delays, and dose reductions on EFS and OS. The analysis is restricted to patients who complete treatment within 180 days since treatment initiation. As discussed at the beginning of the chapter, including patients with less than 6 cycles may not be correct since they may have been treated further but outside the trial. Since it is the intended therapy what we defined as complete therapy in the beginning of the chapter, we are interested in knowing the effects of different therapy modifications under complete therapy.

The analyses have some limitations. For example, short and long cycle delays, as well as small and large dose reductions, are treated equally. Clinicians do not believe that a cycle delay of 3 days could have an effect on patients' survival. Unfortunately, we are not able to estimate the effect of larger delays, since the latter rarely occur. Furthermore, we often define occurrence of a given type of toxicity when the corresponding is assigned a CTCAE grade of 2 or larger, and for some even grade 1 or larger. At the same time, oncologists argue that action is needed only in case of toxicity grade 3 or 4. However, we are only able to estimate the effect of such a severe oral toxicity.

The major limitation of the analysis is that the validity of the results is conditional on correct model specification of the three models for exposure allocation, given exposure and covariate history, and the MSM itself. However, this is an untestable assumption. The only established indicator of correct model specification is mean of stabilised weights equal to 1. We achieve this by using specification 4 in Table 5.6 when modelling cycle delays, and got very close with specification 3 in Table 5.7 as a model for dose reduction, and Model (5.18) for surgery delay. The final weights used in the EFS model have a mean equal to 0.996.



# Chapter 6

## Discussion

### 6.1 Overview of the results and their limitations

The novelty of this thesis consists of adapting the methodology for causal inference through MSMs with IPTW estimator to the complex set-up of data from a randomised controlled trial in chemotherapy for osteosarcoma. Due to the complexity of the data and the research questions, we build models for the joint effect of three or more exposures. Furthermore, these exposures are a mix of longitudinal and point exposures. In the sequel a summary of the results coming from this study is presented.

Chapter 4 presents three models for HRe with increasing complexity. Using the final model, which incorporates all exposures we are interested in we found the following results. Surgery delay of 7 or more days has a strong negative effect on HRe. Since many surgeries were delayed due to administrative reasons, we found no indication that the effect of surgery delay on HRe is confounded by the toxicities experienced during the last preoperative cycle. This study might suggest that closer collaboration between surgeons and oncologists is required for optimal disease management. Every additional delayed cycle is estimated to cause a 2.567 times relative increase in the probability of a GR. This result indicates that the 18-day recovery period between chemotherapy cycles are probably not sufficient for children's body recovery.

Cycles with reduced dose also have a positive causal effect on the probability of GR, although this effect is not statistically significantly different from zero. If validated and supported by medical consideration, these results could be used as a motivation for a new randomised controlled trial to determine the optimal drugs dosage. Additionally, we estimated the effect of an extra preoperative cycle. In spite of the clinical intuition behind the (supposedly) beneficial effect of an additional chemotherapy cycle, its causal effect estimate on GR is negligibly small.

Chapter 5 shows how to build a MSCPHM in order to assess the causal effects of three exposures on EFS and OS using complex longitudinal chemotherapy data. The results for EFS are consistent with those for OS for all exposures but surgery delay. Postponed tumour resection increases the chance of progression, recurrence, and/or metastases, and decreases the risk of death. Yet, neither of the two effect estimates is significantly

different from zero.

We found that short cycle delays (of 3 or more days) increase the chance of an instantaneous adverse event of any kind, and that dose reductions have a protective effect on adverse events. This finding suggest that smaller doses might be enough for the means of the therapy while preventing the patient from death due to excessive toxicity.

All models should be correctly specified for our inference to be valid. Unfortunately, there is no way to test if a model is correct. Yet, we may study the validity of our results to different model specifications. For this purpose we can use the information presented in Chapters 4 and 5.

A comparison of the results in Tables 5.6, 5.7, and 5.8 could serve as a form of sensitivity analysis. In the model for allocation of cycle delay the introduction of non-linear function of age almost doubles the estimate of cycle delay effect on EFS on the log-hazard scale. The magnitude of the standard error is preserved and the change is approximately within 30% of the standard error. Although in absolute terms the change is big, in the context of the uncertainty of this estimate there is no practical difference. However, these results are obtained when the sample is weighted only by the cycle delay component of the weights. Table 5.8 lists a small positive effect of cycle delay on the hazard of an EFS event, contrary to the previously found small negative effect (see Table 5.6). This means that the different components of the weights interfere with each other. Nevertheless, the two effects are small and not statistically significantly different from zero.

A counter-example is dose reduction. Its estimated effect on EFS is statistically significant and is preserved after additional weighting to correct for surgery delay (see Tables 5.7 and 5.8). However, the effect of dose reduction on EFS is much weaker if we use specification 1 or 2 (Table 5.7). Yet, we see that the mean of the weights under specifications 1 and 2 is further away from 1, therefore these models are not correct. We conclude that in case of a strong signal, it is propagated irrespective of the other components of the weights, while weak signals may even change signs.

Correct model specification assumption receives a lot of attention in the literature. Lefebvre *et al.* [42] claim that the models for exposure allocation should not include pure predictors of the treatment assignment. Rather only risk factors for the outcome and confounders of the treatment effect on the outcome should be used. Mortimer *et al.* [43] show that data-driven models for the denominator of the weights improve the efficiency of the IPTW estimator, i.e. the confidence intervals of the effect estimates will be narrower. They give an example where even if the treatment assignment mechanism is known with respect to gender, there might still be residual (random) confounding by age. Therefore, age should take part in the treatment allocation model. An alternative data-driven idea developed by Pullenayegum *et al.* [44] is that the model for the weights should be chosen such that it achieves balance in the confounders between treated and untreated in the weighted sample. They propose estimating covariate-treatment association in the reweighed data set using different models for the weights, and choosing the model which leaves out the least covariate-treatment association. However, the MSM should also be correctly specified.

In the analyses in this thesis we were interested in testing specific hypothesis which defined the functional form of our MSMs. In other cases the information criterion for MSMs developed by Platt et. al [45] could be used to guide the building of the MSM.

## 6.2 Study limitations due to the available data

In this section we discuss the model limitations driven by the features of the available data. We propose alternative solutions that could be explored in future research. This discussion is divided in two parts. Section 6.2.1 focuses on the confounders, and Section 6.2.2 on the exposures.

### 6.2.1 Limitations due to the confounders

The models for allocation of dose reductions and cycle delays use leucopenia and thrombocytopenia CTCAE grades assigned to each patient based on the laboratory measurements of WBC and platelets count at the end of each cycle. The form in which these values are recorded (see Figure 3.2) requires specification of the date of measurement. Unfortunately, the measurements that should have been taken on day 22 of each cycle frequently were taken earlier (see Figure 3.15b). However, our modelling approach relies on the assumption that irrespective of the date of measurement, the recorded values were used to justify the observed therapy modifications. This assumption is more plausible for cycle delays, since they take place right after the end of the previous cycle. Conversely, dose reductions could be decided upon when the cycle starts, which happens a number of days later than the day of the laboratory measurements. At the last day of the delay, another set of measurements should be taken. Based on the values of these measurements three decisions could be taken: to allocate another cycle delay, to reduce the dose, or to allow a cycle to begin without dose reduction. If the first option is taken, we have unmeasured confounder because we do not have access to intermediate blood test results. However, we model cycle delays in such a way that we do not distinguish between a delay of 3 days, followed by another delay of 3 days, where the latter is based on laboratory measurements at day 2 or 3 of the first delay, and a delay of 6 days. Therefore, we were able to overcome this potential unmeasured confounding. If the blood test during a cycle delay results in prescription of dose reduction, we have unmeasured confounding. Despite this fact, the decision to reduce the dose could also have been based on the toxicities experienced during the previous cycle since oral, oto, and cardiac toxicity also require dose reduction or drug discontinuation. Further, if a dose reduction is allocated due to myelosuppression and infection, this could also be based on the measurements that were recorded on the form.

Here it is important to note that: 1) most cycles with reduced dose do not provide full dose due to the carry-over effect, i.e. if a dose was reduced, the reduction should be preserved for later cycles; 2) we constructed the dose reduction indicator in such a way that it does not distinguish between a cycle with 80% dose because the dose was reduced during the previous cycle, and a cycle with 60% dose preceded by an 80%-dose

cycle, i.e. when a secondary dose reduction was applied due to haematological toxicity (could be unmeasured in some cases). Although we identified some potential unmeasured confounding, we believe that its effect is negligible. If the blood test during a cycle delay results in a decision to begin the next cycle with full dose, its results do not indicate haematological toxicity and are recorded as measurements before start of cycle. In fact, the dose might be reduced without allocation of delay. In such cases, all causes of this therapy modification are available.

We have discussed different scenarios of allocation of cycle delays and dose reductions, and we have found that most of the symptoms that could trigger these therapy modifications were measured. However, we still face the problem of inaccurate timing of the available measurements. A way to overcome the limitation that an early blood test introduces, is to fit spline functions for the WBC and platelets counts based on the three measurements within each cycle. These functions could be used to predict the results of the laboratory measurements at day 22 of each cycle. This approach could be implemented in a future research.

Expert opinion suggests that an action in the form of therapy modification is needed only when a toxicity grade of 3 or 4 is observed. However, such severe toxicities are hardly recorded for the patients who took part of this study. At the same time, therapy modifications in terms of cycle delays and dose reductions could be identified for much larger proportion of the patient-cycles than the observed severe toxicities. This suggests that there are differences in the prescribing patterns between the treating oncologists and the designers of the study. Due to the lack of severe toxicities, we use a cut-off value lower than 3 to label the presence of toxicity of a given type. From the statistical point of view this dichotomisation could be regarded as an action that reduces the variation in the data, and respectively the information that it carries. In a future research it might be better to use toxicity grades as continuous variable and model different non-linear functions, since the latter are more biologically plausible, or use more than two categories to allow for difference in oncologists' prescribing preferences. The study in this thesis is based on very limited sample size (see Figure 3.16), which does not allow the use of other functional forms than a linear function or dichotomisation of toxicity grades. Given that the linear function is restrictive and does not suit the way clinicians perceive toxicities, we have decided to dichotomise the side-effects into high-enough and not-high-enough to justify therapy modification.

In the year 2007, I. J. Lewis and colleagues [28] published the results based on the trial. According to the authors, a cycle is delayed if it starts 3 or more days since the planned date, and the dose of each drug is reduced if the given dose is 80% or less of the nominal dose. To enable comparison, we defined the exposures in the same way. In Table 4 the authors list the number of patient-cycles that are delayed or their dose reduced and the corresponding reasons. The authors claim that the reason for cycle delay is unknown for half of the delayed cycles. Although how they reach this conclusion is not completely clear, since the field where the reason for therapy modification is stated is common for both delays and reductions, this could serve as a motivation for sensitivity analysis for unmeasured confounding. Such an analysis falls beyond the scope of this thesis, since

some methodological work is required before it could be performed.

In 1999 Robins [33] introduced the methodology for performing sensitivity analysis for unmeasured confounding in the context of MSMs. This method, in contrast to other existing methods for sensitivity analysis for unmeasured confounding [34, 35, 36, 37, 38], parametrises the bias in the outcome that could result from missing a confounder, rather than the bias in the model parameters. Later Brumback *et al.* [39] extended the methodology to account for repeated measurements. More recently the framework was adapted to allow sensitivity analysis of a Cox proportional hazards model [40]. However, if we intend to perform a sensitivity analysis for unmeasured confounding in this context, we should further extend the methodology to allow for multiple longitudinal and point exposures for both binary and survival outcomes.

### 6.2.2 Limitations due to the exposures

In Section 3.3 we discuss that the cut-off value of 3 days for cycle delay. This choice is based on two arguments. First, we consider important to use definitions consistent with previous published analyses of this dataset. Second, if we use a larger cut-off value, we will end up with less exposed cycles. The latter will diminish our ability to predict the allocation of delays. Using the cut-off of 3 days, there are 70 delayed cycles (see Table 4.7). In the model for allocation of cycle delays we estimate the effects of 6 variables. A well known rule of thumb for reliability of the parameter estimates of logistic regression model is that we need at least 10 occurrences of the state of interest (10 events) per variable. If we are to consider a cycle as delayed only when it is delayed by a week or more, this rule will be violated.

Classifying cycles as delayed and on-time, allows us to estimate the effect of an additional delayed cycle on HRe and patients' survival. If on the contrary we model the number of days between two consecutive cycles, the cumulative exposure (the sum of the delays in days) will not distinguish between a delay of 6 days and three delays of 2 days each. The effects of these two delays are expected to be very different.

A disadvantage of the way dose reductions are modelled is that the indicator of dose reduction does not distinguish between a small dose reduction of 20% for example, and a drug discontinuation, i.e. 100% reduction. Alternatively, received dose intensity could be used, which is a measure that takes into account both timing and dosage of the drugs. It is a ratio of standardised dose, i.e. dose received divided by dose planned, and standardised time, i.e. a cycle length in days divided by the length of a cycle according to the protocol. Received dose intensity of less than 1 indicates that either a delay was applied or the dose was reduced. An advantage of this approach is that the two forms of therapy modification are combined into one measure. However, this comes at a price of interpretability, since the two effects cannot be disentangled.

The flow-chart on Figure 3.16 shows that the records of 17 patients were not used due to missing date of surgery. In Section 3.3 we explain how we use the date of surgery in order to calculate cycle durations and derive cycle delays. Future research might try to use the date of analysis of the resected specimen to estimate the date of surgery in subjects with missing date of surgery.

### 6.3 Data hierarchy

The data under study characterises with a hierarchical structure. The 177 patients eligible for analysis of HRe were treated in 38 hospitals in 8 countries. The distribution across countries is very heterogeneous. For example, 101 patients come from the UK, and only one from Canada. The same applies for the distribution of patients across hospitals. One hospital contributes 43 patients to the sample, other 15 hospitals treated only one study participant each, and 30 out of the all 38 hospitals treated less than 6 patients each. We used Chi-squared test to test for differences in prescription preferences for delays and dose reductions across hospitals and regions. All four test (combinations of two exposures and two types of clusters) return warning about possible inaccurate estimates which is a result of the extremely unequal distribution of patients across clusters. In order to balance the distribution across clusters, we grouped the patients from The Netherlands (28) and Belgium (6) into a Benelux group (34); Argentina (16) and Chile (4) into South America (20); Canada (1) and Denmark (7) into Other (8); we let the UK (101) and Saudi Arabia (14) form independent groups. Cluster size in terms of patients is listed in brackets. Using this new region variable (Benelux, South America, UK, Saudi Arabia, and Other) we repeated the Chi-square test and found that there are statistically significant differences in the allocation of dose reductions. More precisely, the difference between South America (0.9% dose reduced patient-cycles) and the expected overall average (23.7% dose reduced patient-cycles) is statistically significant ( $p$ -value  $< 0.001$ ). Driven by these findings, we introduced a region random effect in the models for allocation of cycle delays and dose reductions. However, the estimate of the variance of the random effect was zero for both models. In the final models presented in Chapters 4 and 5, we do not include random region effect. Further discussion on the problem with data hierarchy within the context of MSMs can be found in [41].

### 6.4 Simulation study

Using the methodology for estimating joint causal effects through MSMs we identified potential collinearity problems if two or more exposures share the same set of confounders. We estimate the probability of occurrence of one of the exposures conditional on the occurrence of the other exposure and a set of confounders. If the exposure on which we condition is very well predicted by the common set of confounders, then multicollinearity will occur. It may have a number of consequences among which unstable parameter estimates, especially the standard errors of parameter estimate. In this setting, we hypothesised that problems with the IPTWs may occur before the signs of multicollinearity. Therefore, we set up a simulation study to investigate our hypothesis.

We simulated two binary toxicity indicators (a toxicity requires therapy modification or not) each taking the value of 1 with probability of 50%. Then we used these data to stochastically assign therapy modifications conditional on the two toxicity indicators. We varied the probability of protocol compliance, i.e. the probability to allocate therapy modification in case any of the toxicities indicates that therapy modification is required,

between 0.6 and 0.99 in steps of 0.05. Full factorial design was performed. We used variance inflation factor  $\geq 10$ , tolerance  $\leq 0.01$ , and condition index  $\geq 30$  as measures of collinearity. Since we simulated no baseline covariates, and a cross-sectional setting, we were able to construct only unstabilised weights. They are bounded below by the value of 1, and can extend to infinity. However, there is no established measure based on which the distribution of the unstabilised weights can be judged. Furthermore, even when we used 99% probability to assign therapy modification if any of the toxicity indicators equals 1, we were not able to reproduce a multicollinearity setting, i.e. none of the collinearity measures exceeded its threshold value.

In order to be able to construct stabilised weights, we performed a second simulation study. The difference from the first one is that we used the empirical joint distribution of patients' age and gender to first simulate these baseline patient characteristics, and then to assign them toxicity grades conditional on age and gender. For the latter we used the empirical distribution of myelosuppression and mucositis of grade 2 or larger conditional on patients' age and gender. However, we did not obtain any different results. We did not observe extreme stabilised weight values, nor mean weight values far from 1. Nevertheless, we still believe that our hypothesis should be further investigated, probably using another motivating example. However, due to time limitations, we did not pursue this opportunity.



# Bibliography

- [1] J. M. Robins, Marginal structural models, in: Proceedings of the American Statistical Association, 1997.
- [2] M. A. Hernán, B. Brumback, and J. M. Robins, Marginal structural models to estimate the joint causal effect of nonrandomised treatments, *Journal of the American Statistical Association*, 2001.
- [3] J. M. Robins, Marginal structural models versus structural nested models as tools for causal inference, in: *Statistical models in epidemiology, the environment, and clinical trials*, Springer, 2000, pp. 95-133.
- [4] J. M. Robins, M. A. Hernán, and B. Brumback, Marginal structural models and causal inference in epidemiology, *Epidemiology* 11 (5) (2000) 550-560.
- [5] S. R. Cole, and M. A. Hernan, Constructing inverse probability weights for marginal structural models, *American journal of epidemiology* 168 (6) (2008) 656-664.
- [6] C. Lancia, C. Spitoni, J. Anninga, and M. Fiocco, Marginal structural models with dose-delay joint-exposure for assessing variations to chemotherapy intensity, to appear in *Statistical Methods in Medical Research* (2016).
- [7] M. A. Hernán, B. Brumback, and J. M. Robins, Marginal structural models to estimate the causal effect of zidovudine on survival of HIV-positive men, *Epidemiology* 11 (5) (2000) 561-570.
- [8] L. A. Cupples, R. B. D'Agostino, K. Anderson, and W. B. Kannel, Comparison of baseline and repeated measure covariate techniques in the Framingham heart study, *Statistics in medicine* 7 (1988) 205-218.
- [9] R. B. D'Agostino, M. Lee, A. J, Belanger, L. A. Cupples, K. Anderson, and W. B. Kannel, Relation of pooled logistic regression to time dependent cox regression analysis: the Framingham heart study, *Statistics in medicine* 9 (1990) 1501-1515.
- [10] W. M. van der Wal, R. B. Geskus, IPW: an R package for inverse probability weighting, *Journal of Statistical Software* 43 (13) (2011) 1-23.
- [11] T. W. Yee, *Vector Generalised Linear and Additive Models With an Implementation in R*, Springer, 2015.

- [12] M. A. Hernán, J. M. Robins, Causal Inference, CRC, (2016), forthcoming.
- [13] P. R. Rosenbaum, The consequences of adjustment for a concomitant variable that has been affected by the treatment, *Journal of the Royal Statistical Society A* (1984) 656-666.
- [14] J. M. Robins, S. Greenland, F. C. Hu, Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association*, 94 (447) (1999), 687-700.
- [15] J. M. Robins, Association, causation, and marginal structural models, *Synthese*, 1999.
- [16] L. M. Bodnar, M. Davidian, A. M. Siega-Riz, and A. A. Tsiatis, Marginal structural models for analysing causal effects of time-dependent treatments: an application in perinatal epidemiology, *American journal of epidemiology* 159 (10) (2004) 926-934.
- [17] M. A. Hernán, B. Brumback, and J. M. Robins, Estimating the causal effect of zidovudine on CD4 count with marginal structural model for repeated measures, *Statistics in medicine* 21 (2002) 1689-1709.
- [18] M. A. Hernán, E. Lanoy, D. Costagliola, and J. M. Robins, Comparison of dynamic treatment regimes via inverse probability weighting, *Basic and Clinical Pharmacology and Toxicology* 98 (2006) 237-242.
- [19] H. R. Banack, and J. S. Kaufman, Estimating the time-varying joint effects of obesity and smoking on all-cause mortality using marginal structural models, *American journal of epidemiology* 168 (2015).
- [20] C. J. Howe, S. R. Cole, S. H. Mehta, and G. D. Kirk, Estimating the effects of multiple time-varying exposures using joint marginal structural models, *Epidemiology* 23 (2012) 574-582.
- [21] J. P. Klein, H. C. Houwelingen, J. G. Ibrahim, and T. H. Scheike, *Handbook of survival analysis*, CRC Press, 2013, Ch. 7, pp. 131-147.
- [22] J. M. Gran, et. al, A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study, *Statistics in medicine*, 29 (26) (2010), 2757-2768.
- [23] J. M. Robins, Correcting for non-compliance in randomized trials using structural nested mean models, *Communications in Statistics-Theory and methods*, 23 (8) (1994), 2379-2412.
- [24] J. Lok, R. Gill, A. Van Der Vaart, J. M. Robins, Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models, *Statistica Neerlandica*, 58 (3) (2004), 271-295.

- [25] J. M. Robins, Structural nested failure time models, *Encyclopaedia of Biostatistics*, (1998).
- [26] J. M. Robins, and M. A. Hernán, *Longitudinal Data Analysis*, CRC Press, 2009, Ch. 23, pp. 553-599.
- [27] <http://www.eortc.be/protoc/details.asp?protocol=80931>
- [28] I. J. Lewis, *et al.*, Improvement in histological response but not survival in osteosarcoma patients treated with intensified chemotherapy: a randomised phase 3 trial of the European Osteosarcoma Intergroup, *Journal of the National Cancer Institute* 99.2 (2007) 112-128.
- [29] I. J. Lewis, S. Weeden, D. Machin, D. Stark, A. W. Craft, Received dose and dose-intensity of chemotherapy and outcome in nonmetastatic extremity osteosarcoma, *Journal of Clinical Oncology* 18 (24) (2000) 4028-4037.
- [30] J. K. Anninga, H. Gelderblom, M. Fiocco, J. R. Kroep, A. H. Taminiau, P. C. Hogendoorn, R. M. Egeler, Chemotherapeutic adjuvant treatment for osteosarcoma: Where do we stand?, *European Journal of Cancer* 47 (16) (2011) 2431-2445.
- [31] U.S. Dept. of Health and Human Services, Common Terminology Criteria for Adverse Events (CTCAE), available online at [http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE\\_4.03\\_2010-06-14\\_QuickReference\\_8.5x11.pdf](http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf).
- [32] M. Foote, The importance of planned dose of chemotherapy on time: do we need to change our clinical practice?, *The Oncologist* 3 (1998) 365-368.
- [33] J. M. Robins, Association, causation, and marginal structural models. *Synthese*, 121(1) (1999) 151-179.
- [34] D. Y. Lin, B. M. Psaty, R. A. Kronmal, Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* (1998) 948-963.
- [35] R. H. Groenwold, D. B. Nelson, K. L. Nichol, A. W. Hoes, E. Hak, Sensitivity analyses to estimate the potential impact of unmeasured confounding in causal research. *International journal of epidemiology* 39(1) (2010) 107-117.
- [36] M. Abrahamowicz, L. M. Bjerre, M. E. Beauchamp, J. LeLorier, R. Burne, The missing cause approach to unmeasured confounding in pharmacoepidemiology, *Statistics in medicine* (2016).
- [37] T. J. VanderWeele, O. A. Arah, Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders, *Epidemiology* (Cambridge, Mass.) 22(1) (2011) 42-52.

- [38] O. A. Arah, Y. Chiba, S. Greenland, Bias formulas for external adjustment and sensitivity analysis of unmeasured confounder, *Annals of epidemiology* 18(8) (2008) 637-646.
- [39] B. A. Brumback, M. A. Hernán, S. J. Haneuse, J. M. Robins, Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures, *Statistics in medicine* 23(5) (2004) 749-767.
- [40] O. Klungsoyr, J. Sexton, I. Sandanger, J. F. Nygård, Sensitivity analysis for unmeasured confounding in a marginal structural Cox proportional hazards model, *Lifetime data analysis* 15(2) (2009) 278-94.
- [41] J. Kasza, K. R. Polkinghorne, M. R. Marshall, S. P. McDonald, R. Wolfe, Clustering and residual confounding in the application of marginal structural models: dialysis modality, vascular access, and mortality. *American journal of epidemiology* 182(6) (2015) 535-543.
- [42] G. Lefebvre, J. A. Delaney, R. W. Platt, Impact of mis-specification of the treatment model on estimates from a marginal structural model, *Statistics in medicine* 27(18) (2008) 3629-3642.
- [43] K. M. Mortimer, R. Neugebauer, M. Van Der Laan, I. B. Tager, An application of model-fitting procedures for marginal structural models. *American Journal of Epidemiology* 162(4) (2005) 382-388.
- [44] E. M. Pullenayegum, C. Lam, C. Manlhiot, B. M. Feldman, Fitting marginal structural models: estimating covariate-treatment associations in the reweighted data set can guide model fitting, *Journal of clinical epidemiology* 61(9) (2008) 875-881.
- [45] R. W. Platt, M. A. Brookhart, S. R. Cole, D. Westreich, E. F. Schisterman, An information criterion for marginal structural models, *Statistics in medicine* 32(8) (2013) 1383-1393.
- [46] D. Rizopoulos, *Joint models for longitudinal and time-to-event data: With applications in R*, CRC Press (2012).

# Appendix A

## R code

### A.1 R code for Chapter 3

```
library(ggplot2)
# Load the data
load("../.. /B0_0x/data/processed_data.RData")

# color-blind-friendly palette
cbPalette <- c("#E69F00", "#56B4E9", "#009E73", "#F0E442", "
  #0072B2", "#D55E00", "#CC79A7", "#000000", "#999999")

#####
# Descriptive statistics: plots
#####

#####
#### Figure 3.3a ####
#####

png(filename="pics_ch3/B0_06_age.png")
p <- ggplot(data = rgstr06, aes(as.numeric(dor - dob)/
  365.25))
p <- p + geom_bar(binwidth = 1, fill = cbPalette[2], color =
  "gray40")
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + xlab(label = "age□(years)") + ylab(label = "
  Frequency")
p <- p + scale_y_continuous(breaks = seq(0 , 25, 5), minor_
  breaks = NULL)
```

```

p <- p + scale_x_continuous(breaks = seq(0, 40, by=5) - 0.5,
  minor_breaks = NULL, labels = seq(0, 40, by=5))
p
dev.off()

#####
#### Figure 3.3b ####
#####

png(filename="pics_ch3/B0_06_gender.png")

perc <- round(table(rgstr06$sex, useNA = "ifany")*100/length
  (rgstr06$sex), 1)
gender <- c("male", "female")

p <- ggplot(data = NULL, aes(x = "", y = as.numeric(perc),
  fill= gender))
p <- p + geom_bar(width = 1, stat = "identity", color='black
  ')
p <- p + coord_polar(theta = "y", start = 0) + ylim(0, 100)
p <- p + guides(fill = guide_legend(override.aes = list(
  colour = NA)))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.title=element_blank(), axis.ticks = element_
  blank())
p <- p + scale_y_continuous(breaks=(cumsum(perc)-perc/2),
  label = perc)
p <- p + scale_fill_manual(values = cbPalette)
p
dev.off()

#####
#### Figure 3.4a ####
#####

library(plotrix)
png(filename="pics_ch3/B0_06_No_cycles.png")
barplot(table(rgstr06$ncyc, useNA = "ifany") - c(rep(0, 5),
  138.4), col = cbPalette[2], axes = F, names.arg = 1:6,
  ylim = c(0, 40), xlab = "number_of_cycles_completed",
  ylab = "Frequency")
axis(2, at = c(seq(0, 40, by = 10)), labels = c(0, 10, 20,
  100, 200))

```

```

axis.break(2, 25)
lines(x = c(6, 7.4) ,y = c(22, 27))
lines(x = c(6, 7.4) ,y = c(22.7, 27.7))
lines(x = c(6, 7.4) ,y = c(22.35, 27.35), col = "white", lwd
      = 4.5)
dev.off()

```

```

#####
#### Figure 3.4b ####
#####

```

```

id <- rgstr06$patid[rgstr06$ncyc != 6]
term <- tapply(folup06$term[folup06$patid %in% id], folup06$
  patid[folup06$patid %in% id], function(x){x[1]})
dat <- data.frame(patid = id, term = term, ncyc = rgstr06$
  ncyc[rgstr06$patid %in% id])

```

```

png(filename="pics_ch3/B0_06_term_trt_by_cycle.png")
reason <- c("Completed", "Disease▯progression", "Excessive▯
  toxicity", "Treatment▯refusal", "Other")[2:5]

```

```

p <- ggplot(data = dat, aes(x = factor(ncyc)))
p <- p + geom_bar(aes(fill = factor(term), labels = reason))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid= element_
  blank())
p <- p + xlab(label = "number▯of▯cycles▯completed") + ylab(
  label = "")
p <- p + scale_fill_manual(values = cbPalette, breaks =
  reason)
p <- p + guides(fill=guide_legend(title="reason"))
p
dev.off()

```

```

#####
#### Figure 3.5a ####
#####

```

```

png(filename="pics_ch3/B0_06_srg_time_in_cycles.png")
perc <- round(table(rgstr06$srg.cyc, useNA = "ifany")*100/
  length(rgstr06$srg.cyc), 1)
labels <- c(levels(rgstr06$srg.cyc), "missing▯surgery▯data")

```

```

p <- ggplot(data = NULL, aes(x = "", y = as.numeric(perc),
  fill= labels))
p <- p + geom_bar(width = 1, stat = "identity", color='black
  ')
p <- p + coord_polar(theta = "y", start = 0) + ylim(0, 100)
p <- p + guides(fill = guide_legend(override.aes = list(
  colour = NA), title = "surgery_performed"))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.title=element_blank(), axis.ticks = element_
  blank())
p <- p + scale_y_continuous(breaks=(cumsum(perc)-perc/2),
  label = perc)
p <- p + scale_fill_manual(values = cbPalette, breaks =
  labels)
p
dev.off()

#####
# calculate SURGERY DELAY
#####

# select patients with dos
id <- rgstr06$patid[is.na(rgstr06$dos) == F]
chemo06DOS <- chemo06[chemo06$patid %in% id, ]
rgstr06DOS <- rgstr06[rgstr06$patid %in% id, ]
folup06DOS <- folup06[folup06$patid %in% id, ]

# last preop cycle indicator
A3 <- NULL
for(i in 1:length(rgstr06DOS$patid)){
  A3 <- c(A3, rep(0, times = as.numeric(rgstr06DOS$ncyc.
    preop)[i] - 1), 1, rep(0, times = (as.numeric(
    rgstr06DOS$ncyc)[i] - as.numeric(rgstr06DOS$ncyc.preop)
    [i])))
}
chemo06DOS$A3 <- A3

rgstr06DOS$del.srg <- as.numeric(rgstr06DOS$dos - chemo06DOS
  $doc[chemo06DOS$A3 == 1] - 21)
##### End

#####
### Figure 3.5b ###

```

```
#####

png(filename="pics_ch3/BO_06_surg_delay.png")
p <- ggplot(data = rgstr06DOS, aes(del.srg))
p <- p + geom_bar(binwidth = 1, fill = cbPalette[2], color =
  "gray40")
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + xlab(label = "day_of_surgery") + ylab(label = "
  Frequency")
p <- p + scale_y_continuous(breaks = seq(0 , 22, 2), minor_
  breaks = NULL)
p <- p + scale_x_continuous(breaks = seq(-10 , 130, 10),
  minor_breaks = NULL, limits = c(-12, 60))
p
dev.off()

#####
#### Figure 3.6a ####
#####

png(filename="pics_ch3/BO_06_HRe.png")

perc <- round(table(rgstr06$resp6, useNA = "ifany")*100/
  length(rgstr06$resp6), 1)
HRe <- c("Good_response", "Bad_response", "Missing")

p <- ggplot(data = NULL, aes(x = "", y = as.numeric(perc),
  fill= HRe))
p <- p + geom_bar(width = 1, stat = "identity", color='black
  ')
p <- p + coord_polar(theta = "y", start = 0) + ylim(0, 100)
p <- p + guides(fill = guide_legend(override.aes = list(
  colour = NA), title="Histological_response"))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.title=element_blank(), axis.ticks = element_
  blank())
p <- p + scale_y_continuous(breaks=(cumsum(perc)-perc/2),
  label = perc)
p <- p + scale_fill_manual(values = cbPalette, breaks = HRe)
p
dev.off()
```

```

#####
#### Figure 3.6b ####
#####
# OS
rgstr06$OS.status <- ifelse(tapply(folup06$surv1, folup06$
  patid, function(x){x[length(x)]}) == 1, 0, 1)
# 0 = censor; 1 = death due to any cause
rgstr06$OS.time <- tapply(folup06$dofu, folup06$patid,
  function(x){x[length(x)]})

# EFS - doprg
# date of first progression or recurrence
rgstr06$doprg <- tapply(folup06$doprg, folup06$patid,
  function(x){x[which(is.na(x) == F)[1]]})
# date of last follow-up visit == rgstr06$OS.time
event.times <- cbind(rgstr06$doprg, rgstr06$OS.time)

rgstr06$EFS.time <- apply(event.times, 1, min, na.rm = TRUE)
which.min.time.doprg <- apply(event.times, 1, which.min)
rgstr06$EFS.status <- ifelse(which.min.time.doprg == 2,
  rgstr06$OS.status, 1) # 1 if event; 0 - censor

# fix survival since randomisation in months
rgstr06$OS.time <- (rgstr06$OS.time - as.numeric(rgstr06$dor
  ))/(365/12)
rgstr06$EFS.time <- (rgstr06$EFS.time - as.numeric(rgstr06$
  dor))/(365/12)

library(survival)

png(filename="pics_ch3/BO_06_KM.png")
OS <- survfit(Surv(OS.time, OS.status)~1, data = rgstr06)
plot(OS, conf.int=F, mark.time=F, col = cbPalette[2], lwd =
  2, xlab = "Months since randomisation", ylab= "Proportion
  surviving", ylim = c(0.3, 1))
EFS <- survfit(Surv(EFS.time, EFS.status)~1, data = rgstr06)
lines(x = c(0, EFS$time), y = c(1, EFS$surv), type = "step",
  col = cbPalette[1], lwd = 2)
legend("topright", legend = c("Overall survival", "Event-
  free survival"), col = cbPalette[2:1], lwd = c(2, 2), bty
  = "n")
dev.off()

```

```

#####
#### Figure 3.7 ####
#####

chemo06$leukNA <- factor(chemo06$leukNA, labels = c(paste("
  grade", 0:4), "missing"), exclude = NULL)
chemo06$neutNA <- factor(chemo06$neutNA, labels = c(paste("
  grade", 0:4), "missing"), exclude = NULL)
chemo06$tromNA <- factor(chemo06$tromNA, labels = c(paste("
  grade", 0:4), "missing"), exclude = NULL)

png(filename="pics_ch3/B0_06_WBC.png")
p <- ggplot(chemo06[, c("cycno", "leukNA")], aes(x = factor(
  cycno)))
p <- p + geom_bar(aes(fill = chemo06$leukNA))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid= element_
  blank())
p <- p + xlab(label = "cycle") + ylab(label = "Frequency")
p <- p + scale_fill_manual(values = cbPalette)
p <- p + guides(fill=guide_legend(title="WBC_toxicity"))
p
dev.off()

png(filename="pics_ch3/B0_06_plt.png")
p <- ggplot(chemo06[, c("cycno", "tromNA")], aes(x = factor(
  cycno)))
p <- p + geom_bar(aes(fill = chemo06$tromNA))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid= element_
  blank())
p <- p + xlab(label = "cycle") + ylab(label = "Frequency")
p <- p + scale_fill_manual(values = cbPalette)
p <- p + guides(fill=guide_legend(title="Platelets_toxicity"
  ))
p
dev.off()

png(filename="pics_ch3/B0_06_neut.png")
p <- ggplot(chemo06[, c("cycno", "neutNA")], aes(x = factor(
  cycno)))
p <- p + geom_bar(aes(fill = chemo06$neutNA))

```

```

p <- p + theme(panel.background = element_rect(fill = "white
"), axis.ticks = element_blank(), panel.grid= element_
blank())
p <- p + xlab(label = "cycle") + ylab(label = "Frequency")
p <- p + scale_fill_manual(values = cbPalette)
p <- p + guides(fill=guide_legend(title="Neutrophils_
toxicity"))
p
dev.off()

png(filename="pics_ch3/B0_06_oral.png")
p <- ggplot(chemo06[, c("cycno", "oral")], aes(x = factor(
cycno)))
p <- p + geom_bar(aes(fill = chemo06$oral))
p <- p + theme(panel.background = element_rect(fill = "white
"), axis.ticks = element_blank(), panel.grid= element_
blank())
p <- p + xlab(label = "cycle") + ylab(label = "Frequency")
p <- p + scale_fill_manual(values = cbPalette)
p <- p + guides(fill=guide_legend(title="Oral_
toxicity"))
p
dev.off()

png(filename="pics_ch3/B0_06_inf.png")
p <- ggplot(chemo06[, c("cycno", "inf")], aes(x = factor(
cycno)))
p <- p + geom_bar(aes(fill = chemo06$inf))
p <- p + theme(panel.background = element_rect(fill = "white
"), axis.ticks = element_blank(), panel.grid= element_
blank())
p <- p + xlab(label = "cycle") + ylab(label = "Frequency")
p <- p + scale_fill_manual(values = cbPalette)
p <- p + guides(fill=guide_legend(title="Infection"))
p
dev.off()

png(filename="pics_ch3/B0_06_oto.png")
p <- ggplot(chemo06[, c("cycno", "oto")], aes(x = factor(
cycno)))
p <- p + geom_bar(aes(fill = chemo06$oto))
p <- p + theme(panel.background = element_rect(fill = "white
"), axis.ticks = element_blank(), panel.grid= element_
blank())

```

```

p <- p + xlab(label = "cycle") + ylab(label = "Frequency")
p <- p + scale_fill_manual(values = cbPalette[c(1:4, 6)])
p <- p + guides(fill=guide_legend(title="Ototoxicity"))
p
dev.off()

png(filename="pics_ch3/B0_06_car.png")
p <- ggplot(chemo06[, c("cycno", "car")], aes(x = factor(
  cycno)))
p <- p + geom_bar(aes(fill = chemo06$car))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid= element_
  blank())
p <- p + xlab(label = "cycle") + ylab(label = "Frequency")
p <- p + scale_fill_manual(values = cbPalette)
p <- p + guides(fill=guide_legend(title="Cardiac toxicity"))
p
dev.off()

```

```

#####
#### Figure 3.8a ####
#####

```

```

png(filename="pics_ch3/B0_06_cyc_dur.png")
p <- ggplot(data = chemo06DOS, aes(x = factor(cycno), y =
  cyc.dur))
p <- p + geom_boxplot()
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + xlab(label = "cycle") + ylab(label = "cycle_
  duration_(days)")
p <- p + scale_y_continuous(breaks = seq(10 , 200, 10),
  minor_breaks = NULL, limits = c(10, 80))
p <- p + geom_hline(aes(yintercept=21), colour = cbPalette
  [1])
p
dev.off()

```

```

#####
#### Figure 3.8b ####
#####

```

```

png(filename="pics_ch3/BO_06_KM_c6.png")
status_c6 <- as.numeric(rgstr06$ncyc == 6)
time_c6 <- tapply(chemo06$doc, chemo06$patid, function(x){x[
  length(x)] - x[1]})
cycle6 <- survfit(Surv(time_c6, status_c6)~1)
plot(cycle6, xlab = "time_since_start_of_cycle_1_(days)",
  ylab = "Proportion_of_patients_yet_to_start_cycle_6",
  mark.time = T, conf.int = F, lwd = 2, col = cbPalette[2])
abline(v = 120, col = cbPalette[1], lwd = 2)
dev.off()

```

```

#####
#### Figure 3.9a ####
#####

```

```

png(filename="pics_ch3/BO_06_delays.png")
delay <- chemo06DOS$cyc.del
delay[which(chemo06DOS$cycno == 1)] <- rgstr06DOS$del.srg

p <- ggplot(data = NULL, aes(x = factor(chemo06DOS$cycno), y
  = delay))
p <- p + geom_boxplot()
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + xlab(label = "") + ylab(label = "delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 40, 5), minor
  _breaks = NULL, limits = c(-11, 42))
p <- p + scale_x_discrete(labels = c("surgery", paste("cycle
  ", 2:6, sep = "_")))
#p <- p + geom_hline(aes(yintercept=21), colour = cbPalette
  [1])
p
dev.off()

```

```

#####
#### Figure 3.9b ####
#####

```

```

png(filename="pics_ch3/BO_06_reas.png")

perc <- round(table(rgstr06DOS$reas, useNA = "ifany")*100/
  length(rgstr06DOS$reas), 1)

```

```

reas <- c("Tumour␣progression", "Haematological␣tox", "Non-
  haematological␣tox", "Administrative", "Other␣reason", "
  Missing")

p <- ggplot(data = NULL, aes(x = "", y = as.numeric(perc),
  fill= reas))
p <- p + geom_bar(width = 1, stat = "identity", color='black
  ')
p <- p + coord_polar(theta = "y", start = 0) + ylim(0, 100)
p <- p + guides(fill = guide_legend(override.aes = list(
  colour = NA), title="Reason␣for␣early/delayed␣surgery"))
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.title=element_blank(), axis.ticks = element_
  blank())
p <- p + scale_y_continuous(breaks=(cumsum(perc)-perc/2),
  label = perc)
p <- p + scale_fill_manual(values = cbPalette, breaks = reas
  )
p
dev.off()

#####
### Figure 3.10 ###
#####

chemo06$sdox <- chemo06$dox/75 # Standardised DOX
chemo06$scddp <- chemo06$cddp/100 # Standardised CDDP
library(grid)
library(gridExtra)
png(filename="pics_ch3/B0_06_chemo_c1_C6.png")
p <- ggplot(chemo06[chemo06$visit == 1, ], aes(x = scddp, y
  = sdox))
p <- p + geom_rect(aes(xmin=0.86, xmax=Inf, ymin=0.86, ymax=
  Inf), fill = "#CCFFCC")
p <- p + geom_point(shape = 1)
p <- p + theme(panel.background = element_rect(fill = "#
  FFF0F0"), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "white"), panel.grid.minor= element
  _line(color = "white"))
p <- p + ggtitle("Cycle␣1")
p <- p + xlab(label = "Standardised␣CDDP␣dose") + ylab(label
  = "Standardised␣DOX␣dose")
p <- p + scale_y_continuous(breaks = seq(0 , 1.4, 0.2),

```

```

  minor_breaks = seq(0 , 1.4, 0.1), limits = c(0, 1.3))
p <- p + scale_x_continuous(breaks = seq(0 , 1.2, 0.2),
  minor_breaks = seq(0 , 1.2, 0.1), limits = c(0, 1.2))
p1 <- p + geom_abline(intercept=0, slope = 1, colour =
  cbPalette[1], linetype = 2)
p1

p <- ggplot(chemo06[chemo06$visit == 2, ], aes(x = scddp, y
  = sdox))
p <- p + geom_rect(aes(xmin=0.86, xmax=Inf, ymin=0.86, ymax=
  Inf), fill = "#CCFFCC")
p <- p + geom_point(shape = 1)
p <- p + theme(panel.background = element_rect(fill = "#
  FFF0F0"), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "white"), panel.grid.minor= element
  _line(color = "white"))
p <- p + ggtitle("Cycle_2")
p <- p + xlab(label = "Standardised_CDDP_dose") + ylab(label
  = "Standardised_DOX_dose")
p <- p + scale_y_continuous(breaks = seq(0 , 1.4, 0.2),
  minor_breaks = seq(0 , 1.4, 0.1), limits = c(0, 1.3))
p <- p + scale_x_continuous(breaks = seq(0 , 1.2, 0.2),
  minor_breaks = seq(0 , 1.2, 0.1), limits = c(0, 1.2))
p2 <- p + geom_abline(intercept=0, slope = 1, colour =
  cbPalette[1], linetype = 2)

p <- ggplot(chemo06[chemo06$visit == 3, ], aes(x = scddp, y
  = sdox))
p <- p + geom_rect(aes(xmin=0.86, xmax=Inf, ymin=0.86, ymax=
  Inf), fill = "#CCFFCC")
p <- p + geom_point(shape = 1)
p <- p + theme(panel.background = element_rect(fill = "#
  FFF0F0"), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "white"), panel.grid.minor= element
  _line(color = "white"))
p <- p + ggtitle("Cycle_3")
p <- p + xlab(label = "Standardised_CDDP_dose") + ylab(label
  = "Standardised_DOX_dose")
p <- p + scale_y_continuous(breaks = seq(0 , 1.4, 0.2),
  minor_breaks = seq(0 , 1.4, 0.1), limits = c(0, 1.3))
p <- p + scale_x_continuous(breaks = seq(0 , 1.2, 0.2),
  minor_breaks = seq(0 , 1.2, 0.1), limits = c(0, 1.2))
p3 <- p + geom_abline(intercept=0, slope = 1, colour =

```

```

cbPalette[1], linetype = 2)

p <- ggplot(chemo06[chemo06$visit == 4, ], aes(x = scddp, y
  = sdox))
p <- p + geom_rect(aes(xmin=0.86, xmax=Inf, ymin=0.86, ymax=
  Inf), fill = "#CCFFCC")
p <- p + geom_point(shape = 1)
p <- p + theme(panel.background = element_rect(fill = "#
  FFF0F0"), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "white"), panel.grid.minor= element
  _line(color = "white"))
p <- p + ggtitle("Cycle_4")
p <- p + xlab(label = "Standardised_CDDP_dose") + ylab(label
  = "Standardised_DOX_dose")
p <- p + scale_y_continuous(breaks = seq(0 , 1.4, 0.2),
  minor_breaks = seq(0 , 1.4, 0.1), limits = c(0, 1.3))
p <- p + scale_x_continuous(breaks = seq(0 , 1.2, 0.2),
  minor_breaks = seq(0 , 1.2, 0.1), limits = c(0, 1.2))
p4 <- p + geom_abline(intercept=0, slope = 1, colour =
  cbPalette[1], linetype = 2)

p <- ggplot(chemo06[chemo06$visit == 5, ], aes(x = scddp, y
  = sdox))
p <- p + geom_rect(aes(xmin=0.86, xmax=Inf, ymin=0.86, ymax=
  Inf), fill = "#CCFFCC")
p <- p + geom_point(shape = 1)
p <- p + theme(panel.background = element_rect(fill = "#
  FFF0F0"), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "white"), panel.grid.minor= element
  _line(color = "white"))
p <- p + ggtitle("Cycle_5")
p <- p + xlab(label = "Standardised_CDDP_dose") + ylab(label
  = "Standardised_DOX_dose")
p <- p + scale_y_continuous(breaks = seq(0 , 1.4, 0.2),
  minor_breaks = seq(0 , 1.4, 0.1), limits = c(0, 1.3))
p <- p + scale_x_continuous(breaks = seq(0 , 1.2, 0.2),
  minor_breaks = seq(0 , 1.2, 0.1), limits = c(0, 1.2))
p5 <- p + geom_abline(intercept=0, slope = 1, colour =
  cbPalette[1], linetype = 2)

p <- ggplot(chemo06[chemo06$visit == 6, ], aes(x = scddp, y
  = sdox))
p <- p + geom_rect(aes(xmin=0.86, xmax=Inf, ymin=0.86, ymax=

```

```

    Inf), fill = "#CCFFCC")
p <- p + geom_point(shape = 1)
p <- p + theme(panel.background = element_rect(fill = "#
  FFF0F0"), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "white"), panel.grid.minor= element
  _line(color = "white"))
p <- p + ggtitle("Cycle_6")
p <- p + xlab(label = "Standardised_CDDP_dose") + ylab(label
  = "Standardised_DOX_dose")
p <- p + scale_y_continuous(breaks = seq(0 , 1.4, 0.2),
  minor_breaks = seq(0 , 1.4, 0.1), limits = c(0, 1.3))
p <- p + scale_x_continuous(breaks = seq(0 , 1.2, 0.2),
  minor_breaks = seq(0 , 1.2, 0.1), limits = c(0, 1.2))
p6 <- p + geom_abline(intercept=0, slope = 1, colour =
  cbPalette[1], linetype = 2)

grid.arrange(p1, p2, p3, p4, p5, p6, ncol=2)
dev.off()

#####
#### Figure 3.12 ####
#####

# exposure -- confounder feedback
# Standardised dose DOX + CDDP
chemo06DOS$st.dose <- 0.5*(chemo06DOS$dox/75 + chemo06DOS$
  cddp/100)

chemo06DOS$inf.num <- as.numeric(chemo06DOS$inf) - 1
chemo06DOS$inf.num[chemo06DOS$inf.num == 5] <- 0

chemo06DOS$oral.num <- as.numeric(chemo06DOS$oral) - 1
chemo06DOS$oral.num[chemo06DOS$oral.num == 5] <- 0

chemo06DOS$oto.num <- as.numeric(chemo06DOS$oto) - 1
chemo06DOS$oto.num[chemo06DOS$oto.num == 5] <- 0

chemo06DOS$car.num <- as.numeric(chemo06DOS$car) - 1
chemo06DOS$car.num[chemo06DOS$car.num == 5] <- 0

chemo06DOS$max.tox <- apply(chemo06DOS[ , c("leuk", "trom",
  "inf.num", "oral.num", "oto.num", "car.num")], 1, max)

```

```

chemo06DOS$which.max.tox <- apply(chemo06DOS[ , c("leuk", "
  trom", "inf.num", "oral.num", "oto.num", "car.num")], 1,
  which.max)
chemo06DOS$which.max.tox <- factor(chemo06DOS$which.max.tox,
  labels = c("WBC", "Plt", "Inf", "Oral", "Oto", "Car"))

chemo06DOS$how.many <- apply(chemo06DOS[ , c("leuk", "trom",
  "inf.num", "oral.num", "oto.num", "car.num")], 1,
  function(x){sum(x >= 2)})

# cyc 2
p <- ggplot(chemo06DOS[chemo06DOS$visit == 2, ], aes(x = st.
  dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 1 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 2]], colour = chemo06DOS$which.max.
  tox[chemo06DOS$visit == 1 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 2]])) + scale_colour
  _manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + ggtitle("Cycle_2")
p <- p + xlab(label = "Stand._dose_DOX+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p2 <- p + guides(colour=FALSE, size=FALSE)

# cyc 3
p <- ggplot(chemo06DOS[chemo06DOS$visit == 3, ], aes(x = st.
  dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 2 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 3]], colour = chemo06DOS$which.max.
  tox[chemo06DOS$visit == 2 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 3]])) + scale_colour
  _manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"), legend.position = c(0.97,

```

```

    0.5))
p <- p + ggtitle("Cycle_3")
p <- p + xlab(label = "Stand. dose_DOX+CDDP") + ylab(label = "Cycle_delay(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor_breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2), minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p3 <- p + guides(colour=FALSE, size=guide_legend(title="CTCAE"))

# cyc 4
p <- ggplot(chemo06DOS[chemo06DOS$visit == 4, ], aes(x = st.dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS$visit == 3 & chemo06DOS$patid %in% chemo06DOS$patid[chemo06DOS$visit == 4]], colour = chemo06DOS$which.max.tox[chemo06DOS$visit == 3 & chemo06DOS$patid %in% chemo06DOS$patid[chemo06DOS$visit == 4]])) + scale_colour_manual(values = cbPalette[c(6:4, 2:1, 7)])

p <- p + theme(panel.background = element_rect(fill = "white"), axis.ticks = element_blank(), panel.grid.major=element_line(color = "gray90"), legend.position = c(0.97, 0.5))
p <- p + ggtitle("Cycle_4")
p <- p + xlab(label = "Stand. dose_DOX+CDDP") + ylab(label = "Cycle_delay(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor_breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2), minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p4 <- p + guides(colour=guide_legend(title="Which_max", override.aes = list(size=5)), size=FALSE)

# cyc 5
p <- ggplot(chemo06DOS[chemo06DOS$visit == 5, ], aes(x = st.dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS$visit == 4 & chemo06DOS$patid %in% chemo06DOS$patid[chemo06DOS$visit == 5]], colour = chemo06DOS$which.max.tox[chemo06DOS$visit == 4 & chemo06DOS$patid %in% chemo06DOS$patid[chemo06DOS$visit == 5]])) + scale_colour

```

```

    _manual(values = cbPalette[c(6:4, 2:1, 7)])

p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + ggtitle("Cycle_5")
p <- p + xlab(label = "Stand._dose_DOX+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p5 <- p + guides(colour=FALSE, size=FALSE)

# cyc 6
p <- ggplot(chemo06DOS[chemo06DOS$visit == 6, ], aes(x = st.
  dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 5 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 6]], colour = chemo06DOS$which.max.
  tox[chemo06DOS$visit == 5 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 6]])) + scale_colour
  _manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + ggtitle("Cycle_6")
p <- p + xlab(label = "Stand._dose_DOX+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p6 <- p + guides(colour=FALSE, size=FALSE)
library(gridExtra)
grid.arrange(p2, p3, p4, p5, p6, ncol=1)

#####
#### Figure 3.13 ####
#####

# cyc 2
p <- ggplot(chemo06DOS[chemo06DOS$visit == 2, ], aes(x = st.

```

```

    dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 1 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 2]], colour = factor(chemo06DOS$how.
  many[chemo06DOS$visit == 1 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 2]])) + scale_
  colour_manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + ggtitle("Cycle_2")
p <- p + xlab(label = "Stand._dose_DOX+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p2 <- p + guides(colour=FALSE, size=FALSE)

# cyc 3
p <- ggplot(chemo06DOS[chemo06DOS$visit == 3, ], aes(x = st.
  dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 2 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 3]], colour = factor(chemo06DOS$how.
  many[chemo06DOS$visit == 2 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 3]])) + scale_
  colour_manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"), legend.position = c(0.97,
  0.5))
p <- p + ggtitle("Cycle_3")
p <- p + xlab(label = "Stand._dose_DOX+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p3 <- p + guides(colour=FALSE, size=guide_legend(title="Max_
  CTCAE"))

# cyc 4

```

```

p <- ggplot(chemo06DOS[chemo06DOS$visit == 4, ], aes(x = st.
  dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 3 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 4]], colour = factor(chemo06DOS$how.
  many[chemo06DOS$visit == 3 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 4]])) + scale_
  colour_manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"), legend.position = c(0.97,
  0.5))
p <- p + ggtitle("Cycle_4")
p <- p + xlab(label = "Stand._dose_DOX_+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p4 <- p + guides(colour=guide_legend(title="How_many_\nCTCAE
  _>=2", override.aes = list(size=5)), size=FALSE)

# cyc 5
p <- ggplot(chemo06DOS[chemo06DOS$visit == 5, ], aes(x = st.
  dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 4 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 5]], colour = factor(chemo06DOS$how.
  many[chemo06DOS$visit == 4 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 5]])) + scale_
  colour_manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + ggtitle("Cycle_5")
p <- p + xlab(label = "Stand._dose_DOX_+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p5 <- p + guides(colour=FALSE, size=FALSE)

```

```

# cyc 6
p <- ggplot(chemo06DOS[chemo06DOS$visit == 6, ], aes(x = st.
  dose, y = cyc.del))
p <- p + geom_point(aes(size = chemo06DOS$max.tox[chemo06DOS
  $visit == 5 & chemo06DOS$patid %in% chemo06DOS$patid[
  chemo06DOS$visit == 6]], colour = factor(chemo06DOS$how.
  many[chemo06DOS$visit == 5 & chemo06DOS$patid %in%
  chemo06DOS$patid[chemo06DOS$visit == 6]])) + scale_
  colour_manual(values = cbPalette[c(6:4, 2:1, 7)])
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + ggtitle("Cycle_6")
p <- p + xlab(label = "Stand._dose_DOX+_CDDP") + ylab(label
  = "Cycle_delay_(days)")
p <- p + scale_y_continuous(breaks = seq(-10 , 20, 5), minor
  _breaks = seq(-10 , 20, 2.5), limits = c(-10, 20))
p <- p + scale_x_continuous(breaks = seq(0.4 , 1.2, 0.2),
  minor_breaks = seq(0.4 , 1.2, 0.1), limits = c(0.4, 1.2))
p6 <- p + guides(colour=FALSE, size=FALSE)
library(gridExtra)
grid.arrange(p2, p3, p4, p5, p6, ncol=1)

#####
#### Figure 3.15a ####
#####

load("../.. /B0_0x/data/raw_data.RData")

png(filename="pics_ch3/B0_06_day_of_surgery_since_doc1.png")
p <- ggplot(data = rgstr06, aes(as.numeric(dos - chemo06$doc
  [chemo06$visit == 1]), fill = time))
p <- p + geom_bar(binwidth = 1)#color = NULL
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(),
  panel.grid.major= element_line(color = "
  gray90"))
p <- p + xlab(label = "day_of_surgery_since_start_of_cycle_1
  _(days)") + ylab(label = "Frequency")
p <- p + scale_y_continuous(breaks = seq(0 , 16, 2), minor_
  breaks = NULL)
p <- p + scale_x_continuous(breaks = seq(0 , 180, 20), minor
  _breaks = NULL, limits = c(20, 100))

```

```

p <- p + guides(fill=guide_legend(title="time_of_surgery"))
p <- p + scale_fill_manual(values = cbPalette)
p
dev.off()

#####
#### Figure 3.15b ####
#####

png(filename="pics_ch3/B0_06_doc_dot3.png")
p <- ggplot(data = chemo06, aes(x = factor(cycno), y = as.
  numeric(dot3 - doc)))
p <- p + geom_boxplot()
p <- p + theme(panel.background = element_rect(fill = "white
  "), axis.ticks = element_blank(), panel.grid.major=
  element_line(color = "gray90"))
p <- p + xlab(label = "cycle") + ylab(label = "day_of_test_
  at_the_end_of_cycle")
p <- p + scale_y_continuous(breaks = seq(-10 , 50, 5), minor
  _breaks = NULL, limits = c(5, 45))
p <- p + geom_hline(aes(yintercept=22), colour = cbPalette
  [1])
p
dev.off()

```

## A.2 R code for Chapter 4

```

library(xtable); library(survey)
#####
#####
##### MODEL 1 #####
#####
#####

load(file = "../B0_0x/data/processed_data.RData")
#select patients with dos and HRe
id <- rgstr06$patid[is.na(rgstr06$dos) == F & is.na(rgstr06$
  resp6) == F]
chemo06HRe <- chemo06[chemo06$patid %in% id, ]
rgstr06HRe <- rgstr06[rgstr06$patid %in% id, ]

# select only patients with two preoperative cycles
# and their preoperative data only

```

```

id <- rgstr06HRe$patid[rgstr06HRe$ncyc.preop == 2]
rgstr06HRe2cyc <- rgstr06HRe[rgstr06HRe$patid %in% id, ]
chemo06HRe2cyc <- chemo06HRe[chemo06HRe$patid %in% id &
  chemo06HRe$cycno %in% c(1, 2), ]

# calculate exposure
chemo06HRe2cyc$A1 <- ifelse(chemo06HRe2cyc$cyc.del >= 3, 1,
  0) # del, not del

# convert toxicities into continuous variables
chemo06HRe2cyc$oral <- as.numeric(chemo06HRe2cyc$oral) - 1
chemo06HRe2cyc$oral[chemo06HRe2cyc$oral == 5] <- 0

# recode outcome variable
rgstr06HRe2cyc$HRe <- ifelse(rgstr06HRe2cyc$resp6 == "good_
  response_<=10pc", 1, 0)

# get from long to wide format
rgstr06HRe2cyc$leuk1 <- chemo06HRe2cyc$leuk[chemo06HRe2cyc$
  cycno == 1]
rgstr06HRe2cyc$leuk2 <- chemo06HRe2cyc$leuk[chemo06HRe2cyc$
  cycno == 2]

rgstr06HRe2cyc$trom1 <- chemo06HRe2cyc$trom[chemo06HRe2cyc$
  cycno == 1]
rgstr06HRe2cyc$trom2 <- chemo06HRe2cyc$trom[chemo06HRe2cyc$
  cycno == 2]

rgstr06HRe2cyc$oral1 <- chemo06HRe2cyc$oral[chemo06HRe2cyc$
  cycno == 1]
rgstr06HRe2cyc$oral2 <- chemo06HRe2cyc$oral[chemo06HRe2cyc$
  cycno == 2]

rgstr06HRe2cyc$inf1 <- chemo06HRe2cyc$inf[chemo06HRe2cyc$
  cycno == 1]
rgstr06HRe2cyc$neur1 <- chemo06HRe2cyc$neur[chemo06HRe2cyc$
  cycno == 1]
rgstr06HRe2cyc$oto1 <- chemo06HRe2cyc$oto[chemo06HRe2cyc$
  cycno == 1]
rgstr06HRe2cyc$car1 <- chemo06HRe2cyc$car[chemo06HRe2cyc$
  cycno == 1]

# center age

```

```

rgstr06HRe2cyc$age16 <- rgstr06HRe2cyc$age - 16

# move cycle 2 delay to registration table
rgstr06HRe2cyc$cyc2del <- chemo06HRe2cyc$A1[chemo06HRe2cyc$
  cycno == 2]

#####
# model: tox -> HRe
tox.HRe <- glm(HRe ~ I(leuk1 >= 2) + I(leuk2 >= 2) + I(oral1
  >= 2) + I(oral2 >= 2) + cyc2del + sex + age16, family =
  binomial, data = rgstr06HRe2cyc)
xtable(summary(tox.HRe), digits = 3)

#####
# model: tox -> exposure
tox.exp <- glm(cyc2del ~ I(leuk1 >= 2) + I(oral1 >= 2) + sex
  + age16, family = binomial, data = rgstr06HRe2cyc)
xtable(summary(tox.exp), digits = 3)

#####
# models exposure -> toxicities
exp.tox1 <- glm(I(leuk2 >= 2) ~ cyc2del + I(leuk1 >= 2) + I(
  oral1 >= 2) + sex + age16, family = binomial, data =
  rgstr06HRe2cyc)

exp.tox2 <- glm(I(oral2 >= 2) ~ cyc2del + I(leuk1 >= 2) + I(
  oral1 >= 2) + sex + age16, family = binomial, data =
  rgstr06HRe2cyc)

#####
# estimation of subject-specific weights
rgstr06HRe2cyc$Wi <- 1/((tox.exp$fitted.values ^
  rgstr06HRe2cyc$cyc2del) * ((1 - tox.exp$fitted.values) ^
  (1 - rgstr06HRe2cyc$cyc2del)))

pres.weights <- apply(cbind(tox.exp$fitted.values, 1/
  rgstr06HRe2cyc$Wi, rgstr06HRe2cyc$Wi), 2, summary)
xtable(t(pres.weights), digits = 3)

#####
# causal model with robust standard errors
MSM1 <- summary(svyglm(HRe ~ cyc2del, design = svydesign(~
  1, weights = ~ Wi, data = rgstr06HRe2cyc), family =

```

```

      quasibinomial()))$coef
M1 <- summary(svyglm(HRe ~ cyc2del, design = svydesign(~ 1,
  weights = ~ 1, data = rgstr06HRe2cyc), family =
  quasibinomial()))$coef
m1.or <- exp(c(MSM1[2, 1], M1[2, 1]))
m1.ci <- exp(c(MSM1[2, 1], M1[2, 1]) + 1.96*matrix(c(-1, 1,
  -1, 1), byrow = T, ncol = 2) * c(MSM1[2, 2], M1[2, 2]))
m1.p <- c(MSM1[2, 4], M1[2, 4])

m1.res <- data.frame(cbind(m1.or, m1.ci, m1.p))
xtable(m1.res, digits = 3)

#####
#####
##### MODEL 2 #####
#####
#####

load(file = "../..//BO_0x/data/processed_data.RData")
library(xtable); library(survey)

#select patients with dos and HRe
id <- rgstr06$patid[is.na(rgstr06$dos) == F & is.na(rgstr06$
  resp6) == F]
chemo06HRe <- chemo06[chemo06$patid %in% id, ]
rgstr06HRe <- rgstr06[rgstr06$patid %in% id, ]

# calculate exposure
chemo06HRe$A1 <- ifelse(chemo06HRe$cyc.del >= 3, 1, 0) #
  delayed, not delayed
chemo06HRe$A2 <- ifelse(0.5*(chemo06HRe$dox/75 + chemo06HRe$
  cddp/100) <= 0.86, 1, 0) # red, no red

# convert toxicities into continuous variables
chemo06HRe$oral <- as.numeric(chemo06HRe$oral) - 1
chemo06HRe$oral[chemo06HRe$oral == 5] <- 0

chemo06HRe$inf <- as.numeric(chemo06HRe$inf) - 1
chemo06HRe$inf[chemo06HRe$inf == 5] <- 0

chemo06HRe$oto <- as.numeric(chemo06HRe$oto) - 1
chemo06HRe$oto[chemo06HRe$oto == 4] <- 0

```

```

chemo06HRe$car <- as.numeric(chemo06HRe$car) - 1
chemo06HRe$car[chemo06HRe$car == 5] <- 0

# recode outcome variable
rgstr06HRe$HRe <- ifelse(rgstr06HRe$resp6 == "good_response_
  <=10pc", 1, 0)

# center age
rgstr06HRe$age16 <- rgstr06HRe$age - 16

# translate age and gender to chemo table
chemo06HRe$age16 <- rep(rgstr06HRe$age - 16, times =
  rgstr06HRe$ncyc)
chemo06HRe$sex <- rep(rgstr06HRe$sex, times = rgstr06HRe$
  ncyc)

# select preoperative chemotherapy data
chemo06HRePreop <- chemo06HRe[chemo06HRe$cyc == "preop", ]

# calculate cumulative toxicity
rgstr06HRe$leuk.cum <- tapply(chemo06HRePreop$leuk,
  chemo06HRePreop$patid, function(x) {sum(x >= 2)})
rgstr06HRe$trom.cum <- tapply(chemo06HRePreop$trom,
  chemo06HRePreop$patid, function(x) {sum(x >= 2)})
rgstr06HRe$inf.cum <- tapply(chemo06HRePreop$inf,
  chemo06HRePreop$patid, function(x) {sum(x >= 2)})
rgstr06HRe$oral.cum <- tapply(chemo06HRePreop$oral,
  chemo06HRePreop$patid, function(x) {sum(x >= 2)})
rgstr06HRe$car.cum <- tapply(chemo06HRePreop$car,
  chemo06HRePreop$patid, function(x) {sum(x >= 1)})
rgstr06HRe$oto.cum <- tapply(chemo06HRePreop$oto,
  chemo06HRePreop$patid, function(x) {sum(x >= 1)})
# calculate cumulative exposure
rgstr06HRe$del.cum <- tapply(chemo06HRePreop$A1,
  chemo06HRePreop$patid, sum)
rgstr06HRe$reduc.cum <- tapply(chemo06HRePreop$A2,
  chemo06HRePreop$patid, sum)

#####
# model: tox -> HRe
tox.risk <- glm(HRe ~ leuk.cum + trom.cum + inf.cum + oral.
  cum + car.cum + oto.cum + del.cum + reduc.cum + ncyc.
  preop + sex + age16, family = binomial, data = rgstr06HRe

```

```

)
xtable(summary(tox.risk), digits = 3)

#####
# shift exposure with one cycle back such that
# exposure in cycle k is available on row k-1
# then a patient with 4 preop cycles contributes with 3 rows
# , i.e
# his/her first three cycles, and exposure shifted one cycle
# back, i.e.
# and exposure for cycles 2 to 4.
chemo06HRePreopk_1 <- chemo06HRePreop[chemo06HRePreop$cycno
  < rep(rgstr06HRe$ncyc.preop, times = rgstr06HRe$ncyc.
    preop), ]

chemo06HRePreopk_1$A1kplus1 <- chemo06HRePreop$A1[
  chemo06HRePreop$cycno > 1]
chemo06HRePreopk_1$A2kplus1 <- chemo06HRePreop$A2[
  chemo06HRePreop$cycno > 1]

#####
# model: tox -> exposure
tox.del <- glm(A1kplus1 ~ I(leuk >= 2) + I(trom >= 2) + I(
  oral >= 2) + A1 + sex + age16, family = binomial, data =
  chemo06HRePreopk_1)
xtable(summary(tox.del), digits = 3)

tox.red <- glm(A2kplus1 ~ I(leuk >= 2 & inf >= 1) + I(trom
  >= 2 & inf >= 1) + I(oto >= 1) + A2 + sex + age16,
  family = binomial, data = chemo06HRePreopk_1)
xtable(summary(tox.red), digits = 3)

# shift toxicities with one cycle back
chemo06HRePreop$leuk_1[chemo06HRePreop$cycno == 1] <- 0
chemo06HRePreop$leuk_1[chemo06HRePreop$cycno > 1] <-
  chemo06HRePreopk_1$leuk

chemo06HRePreop$trom_1[chemo06HRePreop$cycno == 1] <- 0
chemo06HRePreop$trom_1[chemo06HRePreop$cycno > 1] <-
  chemo06HRePreopk_1$trom

chemo06HRePreop$oral_1[chemo06HRePreop$cycno == 1] <- 0
chemo06HRePreop$oral_1[chemo06HRePreop$cycno > 1] <-

```

```

chemo06HRePreopk_1$oral

chemo06HRePreop$inf_1[chemo06HRePreop$cycno == 1] <- 0
chemo06HRePreop$inf_1[chemo06HRePreop$cycno > 1] <-
  chemo06HRePreopk_1$inf

chemo06HRePreop$oto_1[chemo06HRePreop$cycno == 1] <- 0
chemo06HRePreop$oto_1[chemo06HRePreop$cycno > 1] <-
  chemo06HRePreopk_1$oto

#####
# models exposure -> toxicities
exp.leuk <- glm(I(leuk >= 2) ~ A1 + A2 + I(leuk_1 >= 2) + I(
  trom_1 >= 2) + I(oral_1 >= 2) + I(inf_1 >= 1) + I(oto_1
  >= 1) + sex + age16, family = binomial, data =
  chemo06HRePreop)
summary(exp.leuk)

exp.trom <- glm(I(trom >= 2) ~ A1 + A2 + I(leuk_1 >= 2) + I(
  trom_1 >= 2) + I(oral_1 >= 2) + I(inf_1 >= 1) + I(oto_1
  >= 1) + sex + age16, family = binomial, data =
  chemo06HRePreop)
summary(exp.trom)

exp.oral <- glm(I(oral >= 2) ~ A1 + I(leuk_1 >= 2) + I(trom_
  1 >= 2) + I(oral_1 >= 2) + I(inf_1 >= 1) + I(oto_1 >= 1)
  + sex + age16, family = binomial, data = chemo06HRePreop
  )
summary(exp.oral)

exp.inf <- glm(I(inf >= 1) ~ A2 + I(leuk_1 >= 2) + I(trom_1
  >= 2) + I(oral_1 >= 2) + I(inf_1 >= 1) + I(oto_1 >= 1) +
  sex + age16, family = binomial, data = chemo06HRePreop)
summary(exp.inf)

exp.oto <- glm(I(oto >= 1) ~ A2 + I(leuk_1 >= 2) + I(trom_1
  >= 2) + I(oral_1 >= 2) + I(inf_1 >= 1) + I(oto_1 >= 1) +
  sex + age16, family = binomial, data = chemo06HRePreop)
summary(exp.oto)

xtable(rbind(summary(exp.leuk)$coef[2:3, ], summary(exp.trom
  )$coef[2:3, ], summary(exp.oral)$coef[2, ], summary(exp.
  inf)$coef[2, ], summary(exp.oto)$coef[2, ]), digits = 3)

```

```

#####
# estimation of subject-specific weights

# SW ^ del

denom.sw.del <- glm(A1kplus1 ~ cycno + A1 + A2 + I(leuk >=
  2) + I(trom >= 2) + I(oral >= 2) + age16 + sex, family =
  binomial(link = "logit"), data = chemo06HRePreopk_1)

denom.sw.del.i.k <- (denom.sw.del$fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - denom.sw.del$fitted.
  values)^(1 - chemo06HRePreopk_1$A1kplus1))

num.sw.del <- glm(A1kplus1 ~ cycno + A1 + A2, family =
  binomial(link = "logit"), data = chemo06HRePreopk_1)

num.sw.del.i.k <- (num.sw.del$fitted.values^chemo06HRePreopk
  _1$A1kplus1)*((1 - num.sw.del$fitted.values)^(1 -
  chemo06HRePreopk_1$A1kplus1))

sw.del.i <- tapply(num.sw.del.i.k/denom.sw.del.i.k, factor(
  chemo06HRePreopk_1$patid), prod)
sw.del.i <- c(sw.del.i[1:55], 1, sw.del.i[56:78], 1, sw.del.
  i[79:175])

# SW ^ red

denom.sw.red <- glm(A2kplus1 ~ cycno + A1kplus1 + A1 + A2 +
  I(leuk >= 2 & inf >= 1) + I(trom >= 2 & inf >= 1) + I(
  oto >= 1) + age16 + sex, family = binomial(link = "logit
  "), data = chemo06HRePreopk_1)

denom.sw.red.i.k <- (denom.sw.red$fitted.values^
  chemo06HRePreopk_1$A2kplus1)*((1 - denom.sw.red$fitted.
  values)^(1 - chemo06HRePreopk_1$A2kplus1))

num.sw.red <- glm(A2kplus1 ~ cycno + A1kplus1 + A1 + A2,
  family = binomial(link = "logit"), data =
  chemo06HRePreopk_1)

num.sw.red.i.k <- (num.sw.red$fitted.values^chemo06HRePreopk

```

```

    _1$A2kplus1)*((1 - num.sw.red$fitted.values)^(1 -
    chemo06HRePreopk_1$A2kplus1))

sw.red.i <- tapply(num.sw.red.i.k/denom.sw.red.i.k, factor(
  chemo06HRePreopk_1$patid), prod)
sw.red.i <- c(sw.red.i[1:55], 1, sw.red.i[56:78], 1, sw.red.
  i[79:175])

#####
# define last preoperative cycle indicator
last <- rep(0, length(chemo06HRePreop$patid))
last[chemo06HRePreop$visit == 1] <- 1
chemo06HRePreop$last <- c(last[-1], 1)

# SW ~ last

denom.sw.last <- glm(last ~ cycno + A1 + A2 + I(leuk >= 2)
  + I(trom >= 2) + I(oral >= 2) + I(inf >= 2) + I(oto >= 1
  ) + age16 + sex, family = binomial(link = "logit"), data
  = chemo06HRePreop)

denom.sw.last.i.k <- (denom.sw.last$fitted.values^
  chemo06HRePreop$last)*((1 - denom.sw.last$fitted.values)
  ^ (1 - chemo06HRePreop$last))

num.sw.last <- glm(last ~ cycno + A1 + A2, family =
  binomial(link = "logit"), data = chemo06HRePreop)

num.sw.last.i.k <- (num.sw.last$fitted.values^
  chemo06HRePreop$last)*((1 - num.sw.last$fitted.values)^(1
  - chemo06HRePreop$last))

sw.last.i <- tapply(num.sw.last.i.k/denom.sw.last.i.k,
  factor(chemo06HRePreop$patid), prod)

m2.sw.i <- sw.del.i * sw.red.i * sw.last.i

#####
# present weights

xtable(rbind(summary(denom.sw.del.i.k), summary(sw.del.i),
  summary(denom.sw.red.i.k), summary(sw.red.i), summary(
  denom.sw.srg.i.k), summary(sw.srg.i), summary(m2.sw.i)),

```

```

digits = 3)

par(mfrow = c(1, 4), mar = c(1, 3, 3, 0.5))
boxplot(log(sw.del.i), main = expression(log(SW[i]^{del})),
  ylim = c(-1.5, 1.5))
boxplot(log(sw.red.i), main = expression(log(SW[i]^{red})),
  ylim = c(-1.5, 1.5))
boxplot(log(sw.last.i), main = expression(log(SW[i]^{ncyc})),
  , ylim = c(-1.5, 1.5))
boxplot(log(m2.sw.i), main = expression(log(SW[i])), ylim =
  c(-1.5, 1.5))

par(mfrow = c(1, 1), c(5.1,4.1,4.1,2.1))

# calculate cumulative exposure
rgstr06HRe$del.cum <- tapply(chemo06HRePreop$A1,
  chemo06HRePreop$patid, sum)
rgstr06HRe$reduc.cum <- tapply(chemo06HRePreop$A2,
  chemo06HRePreop$patid, sum)

#####
# causal model with robust standard errors
MSM2 <- summary(svyglm(HRe ~ del.cum + reduc.cum + ncyc.
  preop, design = svydesign(~ 1, weights = ~ m2.sw.i, data
  = rgstr06HRe), family = quasibinomial()))$coef
M2 <- summary(svyglm(HRe ~ del.cum + reduc.cum + ncyc.preop,
  design = svydesign(~ 1, weights = ~ 1, data = rgstr06HRe
  ), family = quasibinomial()))$coef
m2.or <- exp(c(MSM2[2:4, 1], M2[2:4, 1]))
m2.ci <- exp(c(MSM2[2:4, 1], M2[2:4, 1]) + 1.96*matrix(c(-1,
  1, -1, 1, -1, 1, -1, 1, -1, 1), byrow = T, ncol =
  2)*c(MSM2[2:4, 2], M2[2:4, 2]))
m2.p <- c(MSM2[2:4, 4], M2[2:4, 4])

m2.res <- data.frame(cbind(m2.or, m2.ci, m2.p))
xtable(m2.res, digits = 3)

#####
#####
##### MODEL 3 #####
#####

```

```

#####
library(xtable); library(survey)

# calculate exposure
rgstr06HRe$del.srg <- as.numeric(rgstr06HRe$dos -
  chemo06HRePreop$doc[chemo06HRePreop$last == 1] - 21)

rgstr06HRe$A3 <- ifelse(rgstr06HRe$del.srg >= 7, 1, 0) # del
, no del

####
# model for surgery delay based on toxicities
tox.srg <- glm(A3 ~ I(chemo06HRePreop$leuk[chemo06HRePreop$
  last == 1] >= 2) + I(chemo06HRePreop$trom[chemo06HRePreop
  $last == 1] >= 2) + I(chemo06HRePreop$oral[
  chemo06HRePreop$last == 1] >= 2) + sex + age, family =
  binomial, data = rgstr06HRe)

xtable(summary(tox.srg), digits = 3)

#####
##### estimate weights #####
#####

#####
# shift exposure with one cycle back such that
# exposure in cycle k is available on row k-1
# then a patient with 4 preop cycles contributes with 3 rows
, i.e
# his/her first three cycles, and exposure shifted one cycle
back, i.e.
# and exposure for cycles 2 to 4.
chemo06HRePreopk_1 <- chemo06HRePreop[chemo06HRePreop$cycno
  < rep(rgstr06HRe$ncyc.preop, times = rgstr06HRe$ncyc.
  preop), ]

chemo06HRePreopk_1$A1kplus1 <- chemo06HRePreop$A1[
  chemo06HRePreop$cycno > 1]
chemo06HRePreopk_1$A2kplus1 <- chemo06HRePreop$A2[
  chemo06HRePreop$cycno > 1]

#####

```

```

##### SW ~ del #####
#####

#####
# model 1: linear terms
m1.denom.sw.del <- glm(A1kplus1 ~ cycno + A1 + A2 + I(leuk
  >= 2) + I(trom >= 2) + I(oral >= 2) + age16 + sex, family
  = binomial(link = "logit"), data = chemo06HRePreopk_1)

m1.denom.sw.del.i.k <- (m1.denom.sw.del$fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m1.denom.sw.del$fitted
  .values)^(1 - chemo06HRePreopk_1$A1kplus1))

m1.num.sw.del <- glm(A1kplus1 ~ cycno + A1 + A2, family =
  binomial(link = "logit"), data = chemo06HRePreopk_1)

m1.num.sw.del.i.k <- (m1.num.sw.del$fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m1.num.sw.del$fitted.
  values)^(1 - chemo06HRePreopk_1$A1kplus1))

m1.sw.del.i <- tapply(m1.num.sw.del.i.k/m1.denom.sw.del.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m1.sw.del.i <- c(m1.sw.del.i[1:55], 1, m1.sw.del.i[56:78],
  1, m1.sw.del.i[79:175])

m1.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
  , design = svydesign(~ 1, weights = ~ m1.sw.del.i, data =
  rgstr06HRe), family = quasibinomial())

#####
# model 2: quadratic cycle term
m2.denom.sw.del <- glm(A1kplus1 ~ cycno + I(cycno^2) + A1 +
  A2 + I(leuk >= 2) + I(trom >= 2) + I(oral >= 2) + age16
  + sex, family = binomial(link = "logit"), data =
  chemo06HRePreopk_1)

m2.denom.sw.del.i.k <- (m2.denom.sw.del$fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m2.denom.sw.del$fitted
  .values)^(1 - chemo06HRePreopk_1$A1kplus1))

m2.num.sw.del <- glm(A1kplus1 ~ cycno + I(cycno^2) + A1 +
  A2, family = binomial(link = "logit"), data =

```

```

chemo06HRePreopk_1)

m2.num.sw.del.i.k <- (m2.num.sw.del.fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m2.num.sw.del.fitted.
  values)^(1 - chemo06HRePreopk_1$A1kplus1))

m2.sw.del.i <- tapply(m2.num.sw.del.i.k/m2.denom.sw.del.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m2.sw.del.i <- c(m2.sw.del.i[1:55], 1, m2.sw.del.i[56:78],
  1, m2.sw.del.i[79:175])

m2.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
  , design = svydesign(~ 1, weights = ~ m2.sw.del.i, data =
  rgstr06HRe), family = quasibinomial())

#####
# model 3: factor(cycle)
m3.denom.sw.del <- glm(A1kplus1 ~ factor(cycno) + A1 + A2 +
  I(leuk >= 2) + I(trom >= 2) + I(oral >= 2) + age16 + sex
  , family = binomial(link = "logit"), data =
  chemo06HRePreopk_1)

m3.denom.sw.del.i.k <- (m3.denom.sw.del.fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m3.denom.sw.del.fitted.
  values)^(1 - chemo06HRePreopk_1$A1kplus1))

m3.num.sw.del <- glm(A1kplus1 ~ factor(cycno) + A1 + A2,
  family = binomial(link = "logit"), data =
  chemo06HRePreopk_1)

m3.num.sw.del.i.k <- (m3.num.sw.del.fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m3.num.sw.del.fitted.
  values)^(1 - chemo06HRePreopk_1$A1kplus1))

m3.sw.del.i <- tapply(m3.num.sw.del.i.k/m3.denom.sw.del.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m3.sw.del.i <- c(m3.sw.del.i[1:55], 1, m3.sw.del.i[56:78],
  1, m3.sw.del.i[79:175])

m3.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
  , design = svydesign(~ 1, weights = ~ m3.sw.del.i, data =

```

```

rgstr06HRe), family = quasibinomial())

library(rms)
#####
# model 4: quadratic cycle term + RCS(3, age16)
m4.denom.sw.del <- glm(A1kplus1 ~ cycno + I(cycno^2) + A1 +
  A2 + I(leuk >= 2) + I(trom >= 2) + I(oral >= 2) + rcs(
  age16, 3) + sex, family = binomial(link = "logit"), data
  = chemo06HRePreopk_1)
summary(m4.denom.sw.del)

m4.denom.sw.del.i.k <- (m4.denom.sw.del$fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m4.denom.sw.del$fitted
  .values)^(1 - chemo06HRePreopk_1$A1kplus1))

m4.sw.del.i <- tapply(m2.num.sw.del.i.k/m4.denom.sw.del.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m4.sw.del.i <- c(m4.sw.del.i[1:55], 1, m4.sw.del.i[56:78],
  1, m4.sw.del.i[79:175])
summary(m4.sw.del.i)
sd(m4.sw.del.i)

m4.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
  , design = svydesign(~ 1, weights = ~ m4.sw.del.i, data =
  rgstr06HRe), family = quasibinomial())
summary(m4.MSM)

#####
# model 5: quadratic cycle term + RCS(3, age16)*gender
m5.denom.sw.del <- glm(A1kplus1 ~ cycno + I(cycno^2) + A1 +
  A2 + I(leuk >= 2) + I(trom >= 2) + I(oral >= 2) + rcs(
  age16, 3)*sex, family = binomial(link = "logit"), data =
  chemo06HRePreopk_1)
summary(m5.denom.sw.del)

m5.denom.sw.del.i.k <- (m5.denom.sw.del$fitted.values^
  chemo06HRePreopk_1$A1kplus1)*((1 - m5.denom.sw.del$fitted
  .values)^(1 - chemo06HRePreopk_1$A1kplus1))

m5.sw.del.i <- tapply(m2.num.sw.del.i.k/m5.denom.sw.del.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m5.sw.del.i <- c(m5.sw.del.i[1:55], 1, m5.sw.del.i[56:78],
  1, m5.sw.del.i[79:175])

```

```

summary(m5.sw.del.i)
sd(m5.sw.del.i)

m5.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
, design = svydesign(~ 1, weights = ~ m5.sw.del.i, data =
rgstr06HRe), family = quasibinomial())
summary(m5.MSM)

#####
# model 6: quadratic cycle term + lin(age16)*gender
m6.denom.sw.del <- glm(A1kplus1 ~ cycno + I(cycno^2) + A1 +
A2 + I(leuk >= 2) + I(trom >= 2) + I(oral >= 2) + age16*
sex, family = binomial(link = "logit"), data =
chemo06HRePreopk_1)
summary(m6.denom.sw.del)

m6.denom.sw.del.i.k <- (m6.denom.sw.del$fitted.values^
chemo06HRePreopk_1$A1kplus1)*((1 - m6.denom.sw.del$fitted
.values)^(1 - chemo06HRePreopk_1$A1kplus1))

m6.sw.del.i <- tapply(m2.num.sw.del.i.k/m6.denom.sw.del.i.k,
factor(chemo06HRePreopk_1$patid), prod)
m6.sw.del.i <- c(m6.sw.del.i[1:55], 1, m6.sw.del.i[56:78],
1, m6.sw.del.i[79:175])
summary(m6.sw.del.i)
sd(m6.sw.del.i)

m6.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
, design = svydesign(~ 1, weights = ~ m6.sw.del.i, data =
rgstr06HRe), family = quasibinomial())
summary(m6.MSM)

#####
##### SW ^ red #####
#####

#####
# model 1: linear cycno and age
m1.denom.sw.red <- glm(A2kplus1 ~ cycno + A1kplus1 + A1 +
A2 + I(leuk >= 2 & inf >= 1) + I(trom >= 2 & inf >= 1) +
I(oto >= 1) + age16 + sex, family = binomial(link = "
logit"), data = chemo06HRePreopk_1)
summary(m1.denom.sw.red)

```

```

m1.denom.sw.red.i.k <- (m1.denom.sw.red$fitted.values^
  chemo06HRePreopk_1$A2kplus1)*((1 - m1.denom.sw.red$fitted
  .values)^(1 - chemo06HRePreopk_1$A2kplus1))

m1.num.sw.red <- glm(A2kplus1 ~ cycno + A1kplus1 + A1 + A2,
  family = binomial(link = "logit"), data =
  chemo06HRePreopk_1)

m1.num.sw.red.i.k <- (m1.num.sw.red$fitted.values^
  chemo06HRePreopk_1$A2kplus1)*((1 - m1.num.sw.red$fitted.
  values)^(1 - chemo06HRePreopk_1$A2kplus1))

m1.sw.red.i <- tapply(m1.num.sw.red.i.k/m1.denom.sw.red.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m1.sw.red.i <- c(m1.sw.red.i[1:55], 1, m1.sw.red.i[56:78],
  1, m1.sw.red.i[79:175])
summary(m1.sw.red.i)
sd(m1.sw.red.i)

m1.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
  , design = svydesign(~ 1, weights = ~ m6.sw.del.i * m1.sw
  .red.i, data = rgstr06HRe), family = quasibinomial())
summary(m1.MSM)

#####
# model 2: lin cycno and age; toxicity separate
m2.denom.sw.red <- glm(A2kplus1 ~ cycno + A1kplus1 + A1 +
  A2 + I(leuk >= 2) + I(inf >= 1) + I(trom >= 2) + I(oto >=
  1 ) + age16 + sex, family = binomial(link = "logit"),
  data = chemo06HRePreopk_1)
summary(m2.denom.sw.red)

m2.denom.sw.red.i.k <- (m2.denom.sw.red$fitted.values^
  chemo06HRePreopk_1$A2kplus1)*((1 - m2.denom.sw.red$fitted
  .values)^(1 - chemo06HRePreopk_1$A2kplus1))

m2.sw.red.i <- tapply(m1.num.sw.red.i.k/m2.denom.sw.red.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m2.sw.red.i <- c(m2.sw.red.i[1:55], 1, m2.sw.red.i[56:78],
  1, m2.sw.red.i[79:175])
summary(m2.sw.red.i)
sd(m2.sw.red.i)

```

```

m2.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
  , design = svydesign(~ 1, weights = ~ m6.sw.del.i * m2.sw
  .red.i, data = rgstr06HRe), family = quasibinomial())
summary(m2.MSM)

#####
# model 8: lin cycno and age; toxicity separate but +
  interaction
m3.denom.sw.red <- glm(A2kplus1 ~ cycno + A1kplus1 + A1 +
  A2 + I(leuk >= 2) * I(inf >= 1) + I(trom >= 2) * I(inf >=
  1) + I(oto >= 1) + age16 + sex, family = binomial(link
  = "logit"), data = chemo06HRePreopk_1)
summary(m3.denom.sw.red)

m3.denom.sw.red.i.k <- (m3.denom.sw.red$fitted.values^
  chemo06HRePreopk_1$A2kplus1)*((1 - m3.denom.sw.red$fitted
  .values)^(1 - chemo06HRePreopk_1$A2kplus1))

m3.sw.red.i <- tapply(m1.num.sw.red.i.k/m3.denom.sw.red.i.k,
  factor(chemo06HRePreopk_1$patid), prod)
m3.sw.red.i <- c(m3.sw.red.i[1:55], 1, m3.sw.red.i[56:78],
  1, m3.sw.red.i[79:175])
summary(m3.sw.red.i)
sd(m3.sw.red.i)

m3.MSM <- svyglm(HRe ~ ncyc.preop + del.cum + reduc.cum + A3
  , design = svydesign(~ 1, weights = ~ m6.sw.del.i * m3.sw
  .red.i, data = rgstr06HRe), family = quasibinomial())
summary(m3.MSM)

#####
##### SW ~ srg.del #####
#####

m1.denom.sw.srg <- glm(A3 ~ ncyc.preop + chemo06HRePreop$
  A1[chemo06HRePreop$last == 1] + chemo06HRePreop$A2[
  chemo06HRePreop$last == 1] + I(chemo06HRePreop$leuk[
  chemo06HRePreop$last == 1] >= 2) + I(chemo06HRePreop$trom
  [chemo06HRePreop$last == 1] >= 2) + I(chemo06HRePreop$
  oral[chemo06HRePreop$last == 1] >= 2) + age16*sex, family
  = binomial, data = rgstr06HRe)
summary(m1.denom.sw.srg)

```

```

m1.denom.sw.srg.i.k <- (m1.denom.sw.srg$fitted.values^
  rgstr06HRe$A3)*((1 - m1.denom.sw.srg$fitted.values)^(1 -
  rgstr06HRe$A3))

m1.num.sw.srg <- glm(A3 ~ nycyc.preop + chemo06HRePreop$A1[
  chemo06HRePreop$last == 1] + chemo06HRePreop$A2[
  chemo06HRePreop$last == 1], family = binomial(link = "
  logit"), data = rgstr06HRe)

m1.num.sw.srg.i.k <- (m1.num.sw.srg$fitted.values^rgstr06HRe
  $A3)*((1 - m1.num.sw.srg$fitted.values)^(1 - rgstr06HRe$
  A3))

m1.sw.srg.i <- tapply(m1.num.sw.srg.i.k/m1.denom.sw.srg.i.k,
  factor(rgstr06HRe$patid), prod)

#####
##### SW ^ cens #####
#####

denom.sw.cens <- glm(last ~ I(cycno == 2) + I(cycno == 3) +
  I(cycno >= 4) + A1 + A2 + I(leuk >= 2) + I(trom >= 2) +
  I(oral >= 2) + I(leuk >= 2 & inf >= 1) + I(trom >= 2 &
  inf >= 1) + I(oto >= 1) + age16 + sex, family = binomial
  (link = "logit"), data = chemo06HRePreop)
summary(denom.sw.cens)

denom.sw.cens.i.k <- (denom.sw.cens$fitted.values^
  chemo06HRePreop$last)*((1 - denom.sw.cens$fitted.values)
  ^ (1 - chemo06HRePreop$last))

num.sw.cens <- glm(last ~ I(cycno == 2) + I(cycno == 3) + I
  (cycno == 4) + I(cycno > 4) + A1 + A2, family = binomial(
  link = "logit"), data = chemo06HRePreop)

num.sw.cens.i.k <- (num.sw.cens$fitted.values^
  chemo06HRePreop$last)*((1 - num.sw.cens$fitted.values)^(1
  - chemo06HRePreop$last))

sw.cens.i <- tapply(num.sw.cens.i.k/denom.sw.cens.i.k,
  factor(chemo06HRePreop$patid), prod)

```

```

m3.sw.i <- m6.sw.del.i * m2.sw.red.i * m1.sw.srg.i * sw.cens
.i

par(mfrow = c(1, 5), mar = c(1, 3, 3, 0.5))
boxplot(log(m6.sw.del.i), main = expression(log(SW[i]^{del}))
), ylim = c(-1.5, 1.5))
boxplot(log(m2.sw.red.i), main = expression(log(SW[i]^{red}))
), ylim = c(-1.5, 1.5))
boxplot(log(m1.sw.srg.i), main = expression(log(SW[i]^{srg}))
), ylim = c(-1.5, 1.5))
boxplot(log(sw.cens.i), main = expression(log(SW[i]^{ncyc}))
), ylim = c(-1.5, 1.5))
boxplot(log(m3.sw.i), main = expression(log(SW[i])), ylim =
c(-1.5, 1.5))

par(mfrow = c(1, 1), c(5.1,4.1,4.1,2.1))

#####
# estimate probability of observed HRe
load(file = "../..//B0_0x/data/processed_data.RData")

rgstr06$resp6 <- factor(rgstr06$resp6, levels = c("good_
response_<=10pc", "bad_response_>10pc", "missing"))
rgstr06$resp6[is.na(rgstr06$resp6)] <- "missing"

rgstr06$country2 <- ifelse(rgstr06$country %in% c("Canada",
"France", "Portugal", "Slovenia", "South_Africa"), "Other
", rgstr06$country)

cens <- glm(rgstr06$resp6 != "missing" ~ age + sex +
country2, family = binomial, data = rgstr06)

W.cens <- 1/cens$fitted.values[rgstr06$resp6 != "missing" &
is.na(rgstr06$dos) == F]

#####
# causal model with robust standard errors
MSM3 <- summary(svyglm(HRe ~ del.cum + reduc.cum + A3 + ncyc
.preop, design = svydesign(~ 1, weights = ~ m3.sw.i*W.
cens, data = rgstr06HRe), family = quasibinomial()))$coef

M3 <- summary(svyglm(HRe ~ del.cum + reduc.cum + A3 + ncyc.
preop, design = svydesign(~ 1, weights = ~ 1, data =

```

```

    rgstr06HRe), family = quasibinomial()))$coef
m3.or <- exp(c(MSM3[2:5, 1], M3[2:5, 1]))
m3.ci <- exp(c(MSM3[2:5, 1], M3[2:5, 1]) + 1.96*matrix(c(-1,
    1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1),
    byrow = T, ncol = 2)*c(MSM3[2:5, 2], M3[2:5, 2]))
m3.p <- c(MSM3[2:5, 4], M3[2:5, 4])

m3.res <- data.frame(cbind(m3.or, m3.ci, m3.p))
xtable(m3.res, digits = 3)

```

### A.3 R code for Chapter 5

```

#####
# SURVIVAL ANALYSIS
#####
load("../..../BO_0x/Data/processed_data.RData")

# remove first death time of patient 49
folup06 <- folup06[-which(folup06$patid == 49 & folup06$
    visit == 3), ]

#select patients with dos
id <- rgstr06$patid[is.na(rgstr06$dos) == F]
chemo06 <- chemo06[chemo06$patid %in% id, ]
rgstr06 <- rgstr06[rgstr06$patid %in% id, ]
folup06 <- folup06[folup06$patid %in% id, ]

# rm patients with reason to terminate == "disease
  progression"
id <- setdiff(rgstr06$patid, folup06$patid[folup06$term == "
    disease┐progression" & is.na(folup06$term) == F])
chemo06 <- chemo06[chemo06$patid %in% id, ]
folup06 <- folup06[folup06$patid %in% id, ]
rgstr06 <- rgstr06[rgstr06$patid %in% id, ]

# rm patients with less than 6 cycles
id <- rgstr06$patid[rgstr06$ncyc == 6]
chemo06 <- chemo06[chemo06$patid %in% id, ]
folup06 <- folup06[folup06$patid %in% id, ]
rgstr06 <- rgstr06[rgstr06$patid %in% id, ]
#####
# OS
#####

```

```

rgstr06$OS.status <- ifelse(tapply(folup06$surv1, folup06$
  patid, function(x){x[length(x)]}) == 1, 0, 1)
# 0 = censor; 1 = death due to any cause
rgstr06$OS.time <- tapply(folup06$dofu, folup06$patid,
  function(x){x[length(x)]})

#####
# EFS - doprg
#####
# date of first progression or recurrence
rgstr06$doprg <- tapply(folup06$doprg, folup06$patid,
  function(x){x[which(is.na(x) == F)[1]]})
# date of last follow-up visit == rgstr06$OS.time
event.times <- cbind(rgstr06$doprg, rgstr06$OS.time)

rgstr06$EFS.time <- apply(event.times, 1, min, na.rm = TRUE)
which.min.time.doprg <- apply(event.times, 1, which.min)
rgstr06$EFS.status <- ifelse(which.min.time.doprg == 2,
  rgstr06$OS.status, 1)
# 1 if event; 0 - censor

# fix survival since chemotherapy initiation in months
rgstr06$OS.time <- (rgstr06$OS.time - as.numeric(tapply(
  chemo06$doc, chemo06$patid, function(x){x[1]})))/(365/12)
rgstr06$EFS.time <- (rgstr06$EFS.time - as.numeric(tapply(
  chemo06$doc, chemo06$patid, function(x){x[1]})))/(365/12)

#####
# compute time of surgery in cycles
z <- numeric(length(rgstr06$patid))
for(i in 1:length(rgstr06$patid)){
  x <- c(as.numeric(chemo06$doc[chemo06$patid == rgstr06$
    patid[i]]), as.numeric(rgstr06$dos[i]) + 1)
  dos <- as.numeric(rgstr06$dos[i])
  y <- x - dos > 0
  z[i] <- which(y == TRUE)[1]
}

rgstr06$srg.cyc <- factor(z, labels = c("between_1_and_
  2", "between_2_and_3", "between_3_and_4", "
  between_4_and_5", "after_6"))

rgstr06$ncyc.preop <- z-1

```

```

#####
# SURGERY DELAY
#####
# last preop cycle indicator
last <- NULL
for(i in 1:length(rgstr06$patid)){
  last <- c(last, rep(0, times = as.numeric(rgstr06$ncyc.
    preop)[i] - 1), 1, rep(0, times = (as.numeric(rgstr06$
    ncyc)[i] - as.numeric(rgstr06$ncyc.preop)[i])))
}
chemo06$lastpreop <- last

rgstr06$del.srg <- as.numeric(rgstr06$dos - chemo06$doc[
  chemo06$lastpreop == 1] - 21)
##### End

#
#####

# num cyc where at least one drug was reduced by at lest 14%
#
#####

#patient 39 => discont DOX due to cardiac tox
chemo06$dox[is.na(chemo06$dox)] <- 0
#patient 145 => discont CDDP, probabaly ??
chemo06$cddp[is.na(chemo06$cddp)] <- 0

chemo06$red.dox <- as.numeric(chemo06$dox/75 <= 0.86)
chemo06$red.cddp <- as.numeric(chemo06$cddp/75 <= 0.86)
chemo06$red <- ifelse(chemo06$red.dox + chemo06$red.cddp >
  0, 1, 0)
rgstr06$ncyc.red <- tapply(chemo06$red, chemo06$patid,
  function(x){sum(x)})

##### End

#####
# num cyc with delay
rgstr06$ncyc.del3d <- tapply(chemo06$cyc.del, chemo06$patid,
  function(x){sum(x >= 3)})

```

```

#####
## landmark
#####

do.complete <- tapply(chemo06$doc, chemo06$patid, function(x
){x[length(x)] + 3 - x[1]})
plot(rgstr06$OS.time ~ do.complete, xlab = "Days to complete
  chemotherapy since its initiation", ylab = "OS (months since
  chemotherapy initiation)")
abline(v = 180, col = 2)
abline(h = 180/(365/12), col = 3)

plot(rgstr06$EFS.time ~ do.complete, xlab = "Days to
  complete chemotherapy since its initiation", ylab = "EFS (
  months since chemotherapy initiation)")
abline(v = 180, col = 2)
abline(h = 180/(365/12), col = 3)

rgstr06_OS180 <- rgstr06[-which(rgstr06$OS.time <= 180/(365/
  12) | do.complete >= 180), ]
rgstr06_EFS180 <- rgstr06[-which(rgstr06$EFS.time <= 180/
  (365/12) | do.complete >= 180), ]

#####
# MSM_data_preprocess
#####

chemo06$A1 <- ifelse(chemo06$cyc.del >= 3, 1, 0) # delayed,
  not delayed
chemo06$A2 <- chemo06$red # dose reduction/ No dose

chemo06$oral <- as.numeric(chemo06$oral) - 1
chemo06$oral[chemo06$oral == 5] <- 0

chemo06$inf <- as.numeric(chemo06$inf) - 1
chemo06$inf[chemo06$inf == 5] <- 0

chemo06$oto <- as.numeric(chemo06$oto) - 1
chemo06$oto[chemo06$oto == 4] <- 0

chemo06$car <- as.numeric(chemo06$car) - 1
chemo06$car[chemo06$car == 5] <- 0

```

```

chemo06$A1kplus1 <- unlist(tapply(chemo06$A1, chemo06$patid,
  function(x){c(x[-1], 0)}))
chemo06$A2kplus1 <- unlist(tapply(chemo06$A2, chemo06$patid,
  function(x){c(x[-1], 0)}))

chemo06$age16 <- rep(rgstr06$age - 16, times = rgstr06$ncyc)
chemo06$sex <- rep(rgstr06$sex, times = rgstr06$ncyc)
chemo06$del.srg <- rep(rgstr06$del.srg, times = rgstr06$ncyc
)

chemo06$L1 <- ifelse(chemo06$leuk >= 2, 1, 0)
chemo06$L2 <- ifelse(chemo06$trom >= 2, 1, 0)
chemo06$L3 <- ifelse(chemo06$oral >= 2, 1, 0)
chemo06$L4 <- ifelse(chemo06$inf >= 2, 1, 0)
chemo06$L5 <- ifelse(chemo06$oto >= 1, 1, 0)
chemo06$L6 <- ifelse(chemo06$car >= 1, 1, 0)

chemo06$F1 <- ifelse(chemo06$leuk >= 2 & chemo06$inf >= 1,
  1, 0)
chemo06$F2 <- ifelse(chemo06$trom >= 2 & chemo06$inf >= 1,
  1, 0)
chemo06$F3 <- ifelse(chemo06$oral >= 3, 1, 0)

chemo06$L1kmin1 <- unlist(tapply(chemo06$L1, chemo06$patid,
  function(x){c(0, x[-length(x)]}))
chemo06$L2kmin1 <- unlist(tapply(chemo06$L2, chemo06$patid,
  function(x){c(0, x[-length(x)]}))
chemo06$L3kmin1 <- unlist(tapply(chemo06$L3, chemo06$patid,
  function(x){c(0, x[-length(x)]}))
chemo06$L4kmin1 <- unlist(tapply(chemo06$L4, chemo06$patid,
  function(x){c(0, x[-length(x)]}))
chemo06$L5kmin1 <- unlist(tapply(chemo06$L5, chemo06$patid,
  function(x){c(0, x[-length(x)]}))
chemo06$L6kmin1 <- unlist(tapply(chemo06$L6, chemo06$patid,
  function(x){c(0, x[-length(x)]}))

chemo06$F1kmin1 <- unlist(tapply(chemo06$F1, chemo06$patid,
  function(x){c(0, x[-length(x)]}))
chemo06$F2kmin1 <- unlist(tapply(chemo06$F2, chemo06$patid,
  function(x){c(0, x[-length(x)]}))
chemo06$F3kmin1 <- unlist(tapply(chemo06$F3, chemo06$patid,
  function(x){c(0, x[-length(x)]}))

```

```

lastcyc <- NULL
for(i in 1:length(rgstr06$patid)){
  lastcyc <- c(lastcyc, rep(0, times = as.numeric(rgstr06$
    nycyc)[i] - 1), 1)
}
chemo06$lastcyc <- lastcyc

chemo06_OS180 <- chemo06[chemo06$patid %in% rgstr06_OS180$
  patid, ]
chemo06_EFS180 <- chemo06[chemo06$patid %in% rgstr06_EFS180$
  patid, ]

rgstr06_OS180$L1 <- tapply(chemo06_OS180$L1, chemo06_OS180$
  patid, sum)
rgstr06_OS180$L2 <- tapply(chemo06_OS180$L2, chemo06_OS180$
  patid, sum)
rgstr06_OS180$L3 <- tapply(chemo06_OS180$L3, chemo06_OS180$
  patid, sum)
rgstr06_OS180$L4 <- tapply(chemo06_OS180$L4, chemo06_OS180$
  patid, sum)
rgstr06_OS180$L5 <- tapply(chemo06_OS180$L5, chemo06_OS180$
  patid, sum)
rgstr06_OS180$L6 <- tapply(chemo06_OS180$L6, chemo06_OS180$
  patid, sum)

rgstr06_OS180$A1 <- tapply(chemo06_OS180$A1, chemo06_OS180$
  patid, sum)
rgstr06_OS180$A2 <- tapply(chemo06_OS180$A2, chemo06_OS180$
  patid, sum)
rgstr06_OS180$A3 <- ifelse(rgstr06_OS180$del.srg >= 7, 1, 0)

rgstr06_OS180$age16 <- rgstr06_OS180$age - 16

rgstr06_EFS180$L1 <- tapply(chemo06_EFS180$L1, chemo06_
  EFS180$patid, sum)
rgstr06_EFS180$L2 <- tapply(chemo06_EFS180$L2, chemo06_
  EFS180$patid, sum)
rgstr06_EFS180$L3 <- tapply(chemo06_EFS180$L3, chemo06_
  EFS180$patid, sum)
rgstr06_EFS180$L4 <- tapply(chemo06_EFS180$L4, chemo06_
  EFS180$patid, sum)
rgstr06_EFS180$L5 <- tapply(chemo06_EFS180$L5, chemo06_
  EFS180$patid, sum)

```

```

rgstr06_EFS180$L6 <- tapply(chemo06_EFS180$L6, chemo06_
  EFS180$patid, sum)

rgstr06_EFS180$A1 <- tapply(chemo06_EFS180$A1, chemo06_
  EFS180$patid, sum)
rgstr06_EFS180$A2 <- tapply(chemo06_EFS180$A2, chemo06_
  EFS180$patid, sum)
rgstr06_EFS180$A3 <- ifelse(rgstr06_EFS180$del.srg >= 7, 1,
  0)

rgstr06_EFS180$age16 <- rgstr06_EFS180$age - 16

chemo06_EFS180_minlastcyc <- chemo06_EFS180[chemo06_EFS180$
  lastcyc == 0, ]
chemo06_OS180_minlastcyc <- chemo06_OS180[chemo06_OS180$
  lastcyc == 0, ]

#####
#####
##### EFS #####
#####
#####
library(survival); library(rms); library(survey);library(
  xtable)

#####
# model: tox -> surv
EFS.tox <- coxph(Surv(EFS.time, EFS.status) ~ L1 + L2 + L3 +
  L4 + L5 + L6 + A1 + A2 + A3 + age16 + sex, data =
  rgstr06_EFS180)
xtable(EFS.tox, digits = 3)

#####
# model: tox -> exposure
del.tox <- glm(A1kplus1 ~ L1 + L2 + L3 + A1 + age16 + sex,
  family = quasibinomial, data = chemo06_EFS180_minlastcyc)
xtable(summary(del.tox), digits = 3)

red.tox <- glm(A2kplus1 ~ F1 + F2 + F3 + L5 + L6 + A2 +sex +
  age16, family = quasibinomial, data = chemo06_EFS180_
  minlastcyc)
xtable(summary(red.tox), digits = 3)

```

```

surg.del.tox <- glm(A3 ~ chemo06_EFS180$L1[chemo06_EFS180$
  lastpreop == 1] + chemo06_EFS180$L2[chemo06_EFS180$
  lastpreop == 1] + chemo06_EFS180$L3[chemo06_EFS180$
  lastpreop == 1] + sex + age16, family = quasibinomial,
  data = rgstr06_EFS180)
xtable(summary(surg.del.tox), digits = 3)

#####
# models exposure - > toxicities

leuk.exp <- glm(L1 ~ A1 + A2 + L1kmin1 + L2kmin1 + L3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])
trom.exp <- glm(L2 ~ A1 + A2 + L1kmin1 + L2kmin1 + L3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])
oral.exp <- glm(L3 ~ A1 + A2 + L1kmin1 + L2kmin1 + L3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])
inf.exp <- glm(L4 ~ A2 + L1kmin1 + L2kmin1 + L3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])
oto.exp <- glm(L5 ~ A2 + L1kmin1 + L2kmin1 + L3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])
car.exp <- glm(L6 ~ A2 + L1kmin1 + L2kmin1 + L3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])

xtable(rbind(summary(leuk.exp)$coef[2:3, ], summary(trom.exp)
  )$coef[2:3, ], summary(oral.exp)$coef[2:3, ], summary(inf
  .exp)$coef[2, ], summary(oto.exp)$coef[2, ], summary(car.
  exp)$coef[2, ]), digits = 3)

F1.exp <- glm(F1 ~ A2 + F1kmin1 + F2kmin1 + F3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno

```

```

> 1, ]
F2.exp <- glm(F2 ~ A2 + F1kmin1 + F2kmin1 + F3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])
F3.exp <- glm(F3 ~ A2 + F1kmin1 + F2kmin1 + F3kmin1 +
  L4kmin1 + L5kmin1 + L6kmin1 + age16 + sex, family =
  quasibinomial, data = chemo06_EFS180[chemo06_EFS180$cycno
  > 1, ])

#####
# estimation of subject-specific weights

#####
##### SW ^ del #####
#####

m1.del.den <- glm(A1kplus1 ~ visit + L1 + L2 + L3 + A1 + A2
  + age16 + sex, family = binomial, data = chemo06_EFS180_
  minlastcyc)
summary(m1.del.den)
m1.del.den.i.k <- (m1.del.den$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m1.del.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m1.del.num <- glm(A1kplus1 ~ visit + A1 + A2, family =
  quasibinomial, data = chemo06_EFS180_minlastcyc)
m1.del.num.i.k <- (m1.del.num$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m1.del.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m1.del.sw <- tapply(m1.del.num.i.k/m1.del.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m1.del.sw)
sd(m1.del.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m1.del.sw))

m2.del.den <- glm(A1kplus1 ~ visit + I(visit^2) + L1 + L2 +
  L3 + A1 + A2 + age16 + sex, family = binomial, data =
  chemo06_EFS180_minlastcyc)
summary(m2.del.den)
m2.del.den.i.k <- (m2.del.den$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m2.del.den$fitted.values)^(1 -

```

```

chemo06_EFS180_minlastcyc$A1kplus1))
m2.del.num <- glm(A1kplus1 ~ visit + I(visit^2) + A1 + A2,
  family = quasibinomial, data = chemo06_EFS180_minlastcyc)
m2.del.num.i.k <- (m2.del.num$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m2.del.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m2.del.sw <- tapply(m2.del.num.i.k/m2.del.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m2.del.sw)
sd(m2.del.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m2.del.sw))

m3.del.den <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1 + L2 + L3 + A1 + A2 +
  age16 + sex, family = binomial, data = chemo06_EFS180_
  minlastcyc)
summary(m3.del.den)
m3.del.den.i.k <- (m3.del.den$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m3.del.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m3.del.num <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1 + A2, family =
  quasibinomial, data = chemo06_EFS180_minlastcyc)
m3.del.num.i.k <- (m3.del.num$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m3.del.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m3.del.sw <- tapply(m3.del.num.i.k/m3.del.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m3.del.sw)
sd(m3.del.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m3.del.sw))

library(rms)
m4.del.den <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1 + L2 + L3 + A1 + A2 +
  rcs(age16, 3) + sex, family = binomial, data = chemo06_
  EFS180_minlastcyc)
summary(m4.del.den)
m4.del.den.i.k <- (m4.del.den$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m4.del.den$fitted.values)^(1 -

```

```

chemo06_EFS180_minlastcyc$A1kplus1))
m4.del.num <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1 + A2, family =
  quasibinomial, data = chemo06_EFS180_minlastcyc)
m4.del.num.i.k <- (m4.del.num$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m4.del.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m4.del.sw <- tapply(m4.del.num.i.k/m4.del.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m4.del.sw)
sd(m4.del.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m4.del.sw))

m5.del.den <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1 + L2 + L3 + A1 + A2 +
  rcs(age16, 3)*sex, family = binomial, data = chemo06_
  EFS180_minlastcyc)
summary(m5.del.den)
m5.del.den.i.k <- (m5.del.den$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m5.del.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m5.del.num <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1 + A2, family =
  quasibinomial, data = chemo06_EFS180_minlastcyc)
m5.del.num.i.k <- (m5.del.num$fitted.values^chemo06_EFS180_
  minlastcyc$A1kplus1)*((1 - m5.del.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A1kplus1))
m5.del.sw <- tapply(m5.del.num.i.k/m5.del.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m5.del.sw)
sd(m5.del.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw))

m6.del.den <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1 + L2 + L3 + A1 + A2 +
  age16*sex + I(age16^2)*sex, family = binomial, data =
  chemo06_EFS180_minlastcyc)
summary(m6.del.den)
m6.del.den.i.k <- (m6.del.den$fitted.values^chemo06_EFS180_

```

```

minlastcyc$A1kplus1)*((1 - m6.del.den$fitted.values)^(1 -
chemo06_EFS180_minlastcyc$A1kplus1))
m6.del.num <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
I(visit == 4) + I(visit == 5) + A1 + A2, family =
quasibinomial, data = chemo06_EFS180_minlastcyc)
m6.del.num.i.k <- (m6.del.num$fitted.values^chemo06_EFS180_
minlastcyc$A1kplus1)*((1 - m6.del.num$fitted.values)^(1 -
chemo06_EFS180_minlastcyc$A1kplus1))
m6.del.sw <- tapply(m6.del.num.i.k/m6.del.den.i.k, chemo06_
EFS180_minlastcyc$patid, prod)
summary(m6.del.sw)
sd(m6.del.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
data = rgstr06_EFS180, weights = m6.del.sw))

#####
##### SW ~ red #####
#####

m1.red.den <- glm(A2kplus1 ~ visit + F1 + F2 + F3 + L5 + L6
+ A1kplus1 + A2 + A1 + age16 + sex, family = binomial,
data = chemo06_EFS180_minlastcyc)
summary(m1.red.den)
m1.red.den.i.k <- (m1.red.den$fitted.values^chemo06_EFS180_
minlastcyc$A2kplus1)*((1 - m1.red.den$fitted.values)^(1 -
chemo06_EFS180_minlastcyc$A2kplus1))
m1.red.num <- glm(A2kplus1 ~ visit + A1kplus1 + A2 + A1,
family = quasibinomial, data = chemo06_EFS180_minlastcyc)
m1.red.num.i.k <- (m1.red.num$fitted.values^chemo06_EFS180_
minlastcyc$A2kplus1)*((1 - m1.red.num$fitted.values)^(1 -
chemo06_EFS180_minlastcyc$A2kplus1))
m1.red.sw <- tapply(m1.red.num.i.k/m1.red.den.i.k, chemo06_
EFS180_minlastcyc$patid, prod)
summary(m1.red.sw)
sd(m1.red.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
data = rgstr06_EFS180, weights = m5.del.sw*m1.red.sw))

m2.red.den <- glm(A2kplus1 ~ visit + I(visit^2) + F1 + F2 +
F3 + L5 + L6 + A1kplus1 + A2 + A1 + age16 + sex, family
= binomial, data = chemo06_EFS180_minlastcyc)
summary(m2.red.den)

```

```

m2.red.den.i.k <- (m2.red.den$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m2.red.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m2.red.num <- glm(A2kplus1 ~ visit + I(visit^2) + A1kplus1
  + A2 + A1, family = quasibinomial, data = chemo06_EFS180_
  minlastcyc)
m2.red.num.i.k <- (m2.red.num$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m2.red.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m2.red.sw <- tapply(m2.red.num.i.k/m2.red.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m2.red.sw)
sd(m2.red.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw*m2.red.sw))

m3.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + F1 + F2 + F3 + L5 + L6 +
  A1kplus1 + A2 + A1 + age16 + sex, family = binomial,
  data = chemo06_EFS180_minlastcyc)
summary(m3.red.den)
m3.red.den.i.k <- (m3.red.den$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m3.red.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m3.red.num <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1kplus1 + A2 + A1,
  family = quasibinomial, data = chemo06_EFS180_minlastcyc)
m3.red.num.i.k <- (m3.red.num$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m3.red.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m3.red.sw <- tapply(m3.red.num.i.k/m3.red.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m3.red.sw)
sd(m3.red.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw*m3.red.sw))

m4.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + F1 + F2 + F3 + L5 + L6 +
  A1kplus1 + A2 + A1 + rcs(age16, 3) + sex, family =
  binomial, data = chemo06_EFS180_minlastcyc)

```

```

summary(m4.red.den)
m4.red.den.i.k <- (m4.red.den$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m4.red.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m4.red.num <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1kplus1 + A2 + A1,
  family = quasibinomial, data = chemo06_EFS180_minlastcyc)
m4.red.num.i.k <- (m4.red.num$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m4.red.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m4.red.sw <- tapply(m4.red.num.i.k/m4.red.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m4.red.sw)
sd(m4.red.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw*m4.red.sw))

m5.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + F1 + F2 + F3 + L5 + L6 +
  A1kplus1 + A2 + A1 + age16 + I(age16^2) + sex, family =
  binomial, data = chemo06_EFS180_minlastcyc)
summary(m5.red.den)
m5.red.den.i.k <- (m5.red.den$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m5.red.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m5.red.num <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1kplus1 + A2 + A1,
  family = quasibinomial, data = chemo06_EFS180_minlastcyc)
m5.red.num.i.k <- (m5.red.num$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m5.red.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m5.red.sw <- tapply(m5.red.num.i.k/m5.red.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m5.red.sw)
sd(m5.red.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw*m5.red.sw))

m6.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1kplus1 + A2 + A1 + F1
  + F2 + F3 + L5 + L6 + age16*sex + I(age16^2)*sex, family

```

```

    = binomial, data = chemo06_EFS180_minlastcyc)
summary(m6.red.den)

m7.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + F1 + F2 + F3 + L5 + L6 +
  A1kplus1 + A2 + A1 + age16 * sex, family = binomial,
  data = chemo06_EFS180_minlastcyc)
summary(m7.red.den)

m8.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1 + L2 + F3 + L4 + L5 +
  L6 + A1kplus1 + A2 + A1 + age16 + sex, family = binomial
  , data = chemo06_EFS180_minlastcyc)
summary(m8.red.den)
m8.red.den.i.k <- (m8.red.den$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m8.red.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m8.red.num <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1kplus1 + A2 + A1,
  family = quasibinomial, data = chemo06_EFS180_minlastcyc)
m8.red.num.i.k <- (m8.red.num$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m8.red.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m8.red.sw <- tapply(m8.red.num.i.k/m8.red.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m8.red.sw)
sd(m8.red.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw*m8.red.sw))

m9.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1*L4 + L2*L4 + F3 + L5
  + L6 + A1kplus1 + A2 + A1 + age16 + sex, family =
  binomial, data = chemo06_EFS180_minlastcyc)
summary(m9.red.den)
m9.red.den.i.k <- (m9.red.den$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m9.red.den$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m9.red.num <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1kplus1 + A2 + A1,

```

```

    family = quasibinomial, data = chemo06_EFS180_minlastcyc)
m9.red.num.i.k <- (m9.red.num$fitted.values^chemo06_EFS180_
  minlastcyc$A2kplus1)*((1 - m9.red.num$fitted.values)^(1 -
  chemo06_EFS180_minlastcyc$A2kplus1))
m9.red.sw <- tapply(m9.red.num.i.k/m9.red.den.i.k, chemo06_
  EFS180_minlastcyc$patid, prod)
summary(m9.red.sw)
sd(m9.red.sw)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw*m9.red.sw))

#####
##### SW ^ srg.del #####
#####

m1.denom.sw.srg <- glm(A3 ~ ncyc.preop + chemo06_EFS180$A1[
  chemo06_EFS180$lastpreop == 1] + chemo06_EFS180$A2[
  chemo06_EFS180$lastpreop == 1] + chemo06_EFS180$L1[
  chemo06_EFS180$lastpreop == 1]+ chemo06_EFS180$L2[chemo06_
  _EFS180$lastpreop == 1]+ chemo06_EFS180$L3[chemo06_EFS180
  $lastpreop == 1] + age16 + I(age16^2) + sex, family =
  binomial, data = rgstr06_EFS180)
summary(m1.denom.sw.srg)
m1.denom.sw.srg.i.k <- (m1.denom.sw.srg$fitted.values^
  rgstr06_EFS180$A3)*((1 - m1.denom.sw.srg$fitted.values)
  ^((1 - rgstr06_EFS180$A3)))
m1.num.sw.srg <- glm(A3 ~ ncyc.preop + chemo06_EFS180$A1[
  chemo06_EFS180$lastpreop == 1] + chemo06_EFS180$A2[
  chemo06_EFS180$lastpreop == 1], family = binomial(link =
  "logit"), data = rgstr06_EFS180)
m1.num.sw.srg.i.k <- (m1.num.sw.srg$fitted.values^rgstr06_
  EFS180$A3)*((1 - m1.num.sw.srg$fitted.values)^(1 -
  rgstr06_EFS180$A3))
m1.sw.srg.i <- tapply(m1.num.sw.srg.i.k/m1.denom.sw.srg.i.k,
  factor(rgstr06_EFS180$patid), prod)
summary(m1.sw.srg.i)
sd(m1.sw.srg.i)
summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 + A3,
  data = rgstr06_EFS180, weights = m5.del.sw*m8.red.sw*m1.
  sw.srg.i))

#####
# SW ^ selection bias #

```

```
#####

rgstr06_EFS180all$excl <- ifelse(rgstr06_EFS180all$ncyc < 6,
  1, 0)
chemo06_EFS180all$excl <- rep(rgstr06_EFS180all$excl, times
  = rgstr06_EFS180all$ncyc)

denom.sw.cens <- glm(excl ~ + A1 + A2 + A3 + ncyc.preop +
  L1 + L2 + L3 + L4 + L5 + L6 + age16 + sex, family =
  binomial(link = "logit"), data = rgstr06_EFS180all)
summary(denom.sw.cens)
denom.sw.cens <- glm(excl ~ age16 + sex, family = binomial(
  link = "logit"), data = rgstr06_EFS180all)
summary(denom.sw.cens)
denom.sw.cens.i.k <- (denom.sw.cens$fitted.values^rgstr06_
  EFS180all$excl)*((1 - denom.sw.cens$fitted.values)^(1 -
  rgstr06_EFS180all$excl))
num.sw.cens <- glm(lastcyc ~ I(cycno < 6) + A1 + A2, family
  = binomial(link = "logit"), data = chemo06_EFS180)
num.sw.cens.i.k <- (num.sw.cens$fitted.values^chemo06_EFS180
  $lastcyc)*((1 - num.sw.cens$fitted.values)^(1 - chemo06_
  EFS180$lastcyc))
sw.cens.i <- tapply(num.sw.cens.i.k/denom.sw.cens.i.k,
  factor(chemo06_EFS180$patid), prod)
summary(sw.cens.i)
summary(m5.del.sw*m8.red.sw*m1.sw.srg.i)

#####
# present weights' distribution

par(mfrow = c(1, 4), mar = c(1, 3, 3, 0.5))
boxplot(log(m5.del.sw), main = expression(log(SW[i]^{del})),
  ylim = c(-3, 2.05))
boxplot(log(m8.red.sw), main = expression(log(SW[i]^{red})),
  ylim = c(-3, 2.05))
boxplot(log(m1.sw.srg.i), main = expression(log(SW[i]^{srg})
  ), ylim = c(-3, 2.05))
boxplot(log(m5.del.sw*m8.red.sw*m1.sw.srg.i), main =
  expression(log(SW[i])), ylim = c(-3, 2.05))
par(mfrow = c(1, 1), c(5.1,4.1,4.1,2.1))

#####
```

```

# causal model with robust standard errors

MSMEFS <- summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2
  + A3, data = rgstr06_EFS180, weights = m3.del.sw*m8.red.
  sw*m1.sw.srg.i))$coef
EFS <- summary(coxph(Surv(EFS.time, EFS.status) ~ A1 + A2 +
  A3, data = rgstr06_EFS180))$coef
m3.or <- exp(c(MSMEFS[, 1], EFS[, 1]))
m3.ci <- exp(c(MSMEFS[, 1], EFS[, 1]) + 1.96*matrix(c(-1,
  1, -1, 1, -1, 1, -1, 1, -1, 1, -1, 1), byrow = T, ncol =
  2)*c(MSMEFS[, 3], EFS[, 3]))
m3.p <- c(MSMEFS[, 5], EFS[, 5])
m3.res <- data.frame(cbind(m3.or, m3.ci, m3.p))
xtable(m3.res, digits = 3)

#####
#####
##### overall survival #####
#####
#####

# SW ^ del
m5.del.den <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1 + L2 + L3 + A1 + A2 +
  rcs(age16, 3)*sex, family = binomial, data = chemo06_
  OS180_minlastcyc)
summary(m5.del.den)
m5.del.den.i.k <- (m5.del.den$fitted.values^chemo06_OS180_
  minlastcyc$A1kplus1)*((1 - m5.del.den$fitted.values)^(1 -
  chemo06_OS180_minlastcyc$A1kplus1))
m5.del.num <- glm(A1kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1 + A2, family =
  quasibinomial, data = chemo06_OS180_minlastcyc)
m5.del.num.i.k <- (m5.del.num$fitted.values^chemo06_OS180_
  minlastcyc$A1kplus1)*((1 - m5.del.num$fitted.values)^(1 -
  chemo06_OS180_minlastcyc$A1kplus1))
m5.del.sw <- tapply(m5.del.num.i.k/m5.del.den.i.k, chemo06_
  OS180_minlastcyc$patid, prod)
summary(m5.del.sw)
sd(m5.del.sw)

```

```

# SW ~ red
m8.red.den <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + L1 + L2 +F3 + L4 + L5 +
  L6 + A1kplus1 + A2 + A1 + age16 + sex, family = binomial,
  data = chemo06_OS180_minlastcyc)
summary(m8.red.den)
m8.red.den.i.k <- (m8.red.den$fitted.values^chemo06_OS180_
  minlastcyc$A2kplus1)*((1 - m8.red.den$fitted.values)^(1 -
  chemo06_OS180_minlastcyc$A2kplus1))
m8.red.num <- glm(A2kplus1 ~ I(visit == 2) + I(visit ==3) +
  I(visit == 4) + I(visit == 5) + A1kplus1 + A2 + A1,
  family = quasibinomial, data = chemo06_OS180_minlastcyc)
m8.red.num.i.k <- (m8.red.num$fitted.values^chemo06_OS180_
  minlastcyc$A2kplus1)*((1 - m8.red.num$fitted.values)^(1 -
  chemo06_OS180_minlastcyc$A2kplus1))
m8.red.sw <- tapply(m8.red.num.i.k/m8.red.den.i.k, chemo06_
  OS180_minlastcyc$patid, prod)
summary(m8.red.sw)
sd(m8.red.sw)

```

```

# SW ~ srg
m1.denom.sw.srg <- glm(A3 ~ ncy.c.preop + chemo06_OS180$A1[
  chemo06_OS180$lastpreop == 1] + chemo06_OS180$A2[chemo06_
  OS180$lastpreop == 1] + chemo06_OS180$L1[chemo06_OS180$
  lastpreop == 1]+ chemo06_OS180$L2[chemo06_OS180$lastpreop
  == 1]+ chemo06_OS180$L3[chemo06_OS180$lastpreop == 1] +
  age16 + I(age16^2) + sex, family = binomial, data =
  rgstr06_OS180)
summary(m1.denom.sw.srg)
m1.denom.sw.srg.i.k <- (m1.denom.sw.srg$fitted.values^
  rgstr06_OS180$A3)*((1 - m1.denom.sw.srg$fitted.values)^(1
  - rgstr06_OS180$A3))
m1.num.sw.srg <- glm(A3 ~ ncy.c.preop + chemo06_OS180$A1[
  chemo06_OS180$lastpreop == 1] + chemo06_OS180$A2[chemo06_
  OS180$lastpreop == 1], family = binomial(link = "logit"),
  data = rgstr06_OS180)
m1.num.sw.srg.i.k <- (m1.num.sw.srg$fitted.values^rgstr06_
  OS180$A3)*((1 - m1.num.sw.srg$fitted.values)^(1 - rgstr06
  _OS180$A3))
m1.sw.srg.i <- tapply(m1.num.sw.srg.i.k/m1.denom.sw.srg.i.k,
  factor(rgstr06_OS180$patid), prod)
summary(m1.sw.srg.i)

```

```

sd(m1.sw.srg.i)

#####
# causal model with robust standard errors
MSMOS <- summary(coxph(Surv(OS.time, OS.status) ~ A1 + A2 +
  A3, data = rgstr06_OS180, weights = m5.del.sw*m8.red.sw*
  m1.sw.srg.i))$coef
OS <- summary(coxph(Surv(OS.time, OS.status) ~ A1 + A2 + A3,
  data = rgstr06_OS180))$coef
m3.or <- exp(c(MSMOS[ , 1], OS[ , 1]))
m3.ci <- exp(c(MSMOS[ , 1], OS[ , 1]) + 1.96*matrix(c(-1, 1,
  -1, 1, -1, 1, -1, 1, -1, 1), byrow = T, ncol = 2)
  *c(MSMOS[ , 3], OS[ , 3]))
m3.p <- c(MSMOS[ , 5], OS[ , 5])
m3.res <- data.frame(cbind(m3.or, m3.ci, m3.p))
xtable(m3.res, digits = 3)

```